DOCUMENT RESUME

ED 293 387	HE 020 550

AUTHOR	McGuinness, Kathleen, Comp.
TITLE	Legislative Action & Assessment: Reason & Reality. Conference Proceedings (Arlington, Virginia, July 7-9, 1986).
INSTITUTION	American Association of State Colleges and Universities, Washington, D.C.; George Mason Univ., Fairfax, Va.
PUB DATE	Jul 86
NOTE	324p.; Report collected as part of the American Association for Higher Education Assessment Forum.
PUB TYPE	Collected Works - Conference Proceedings (021)
EDRS PRICE	MF01/PC13 Plus Postage.
DESCRIPTORS	*Educational Assessment; Educational Quality; *Higher Education; *Outcomes of Education; *Program Evaluation; Public Colleges; *Public Policy; *Student Evaluation
IDENTIFIERS	*AAHE Assessment Forum; College Outcomes Assessment; Value Added

ABSTRACT

Legislative action and assessment in higher education are addressed in 24 papers from a 1986 conference sponsored by George Mason University (Virginia) and the American Association of State Colleges and Universities. Titles and authors include: "Defining and Assessing Baccalaureate Skills: Ten Case Studies" (American Association of State Colleges and Universities); "Assessing Outcomes: The Real Value Added Is in the Process" (T.W. Banta, H.S. Fisher); "Assessment and the Reification of Learning" (C. Bouton); "The Relationship between High School Writing Experiences and University Admission Requirements" (J.E. Creeden); "Academic Assessment: Implications for Faculty Evaluation" (R. DeBard); "Clayton State College" (J. Doig, L. Greer, J. Swofford); "Performance Funding for Tennessee Higher Education: How Does It Work and Where Did It Come From" (R.G. Dumont); "Stimulating Local Assessment Initiatives by State Policy: Miami University's Response to Ohio Program Excellence" (C.W. Ellison); "The Assessment of Academic Quality: A Review of Existing Techniques with Emphasis on Their Validity and Feasibility in Public Higher Education" (W.F. Enteman, P.I. Jackson); "Where Is the Value in "Value-Added" Testing?" (H.T. Everson); "An Approach to 'In-House' Test Development: For Major Field Assessment" (A.J. Golden); "Assessment in Undergraduate Education and Some Implications for Minorities" (F.A. Hrabowski); "Miami-Dade Community College" (J.T. Kelly, J. Losak); "Evidence of Quality Assessment and Control in State Budget Documents" (J.A. Lamb); "Committee on Student Assessment" (F. Luth, V. Livingston, A. Kohen, D. Corbin, A. Lyndrup); "Creating Usable Program Evaluation" (J.H. McBath); "State Control of Higher Education in Michigan: A New Scenario" (T.H. McKinney); "An Overview of Institutional Evaluation" (R.I. Miller); "The Role of Professional Judgment in the Use of Assessment Data" (E.T. Moran); "The Promise of Assessment: Linking Indicators to Outcomes" (T.E.R. Redmon); "Evaluation of Excellence: The Tennessee Plan" (L. Rudolph, J. Nixon); "Teaching Thinking: Implications for Institutional and Public Policy" (M.A. Schlessinger); "The Uses of Assessment for Decision Making: A Primer About Northeast Missouri State University's Value-Added Program" (T.B. Smith); and "Evaluating Academic Programs" (H.L. Thompson). (SW)

ED293387

HE 020 550

George Mason University

Proceedings from the Conference on Legislative Action and Assessment: Reason and Reality

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
George Mason
University
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

July 7-9, 1986
Hyatt Arlington, Arlington, Virginia

One Dupont Circle
Suite 600
Washington, D.C. 20036
202/293-6440

AAHE
AMERICAN ASSOCIATION
FOR HIGHER EDUCATION

The AAHE ASSESSMENT FORUM is a three-year project supported by the Fund for the Improvement of Postsecondary Education. It entails three distinct but overlapping activities:

--an annual conference
(the first scheduled for June 14-17, 1987, in Denver)

--commissioned papers
(focused on implementation and other timely assessment concerns; available through the Forum for a small fee)

--information services
(including consultation, referrals, a national directory, and more)

This paper is part of an on-going assessment collection maintained by the Forum. We are pleased to make it more widely available through the ERIC system.

For further information about ASSESSMENT FORUM activities, contact Patricia Hutchings, Director, AAHE ASSESSMENT FORUM, One Dupont Circle, Suite 600, Washington, DC 20036

Board of Directors

Chair
Joseph F. Kauffman
University of Wisconsin
Madison

Chair-Elect
Adele S. Simmons
Hampshire College

Vice Chair
Reatha Clark King
Metropolitan State University

Past Chair
Harriet W. Sheridan
Brown University

Carlos H. Arce
NuStats, Inc.

Estela M. Bensimon
Teachers College
Columbia University

Anne L. Bryant
American Association of
University Women

Donald L. Fruehling
McGraw-Hill, Inc.

Ellen V. Futter
Barnard College

Jerry G. Gaff
Hamline University

Zelda F. Gamson
University of Michigan

Stephen R. Graubard
Daedalus

Joseph Katz
State University of New York
at Stony Brook

Arthur E. Levine
Bradford College

Frank Newman
Education Commission
of the States

Alan Pifer
Carnegie Corporation
of New York

W. Ann Reynolds
The California State
University

Piedad F. Robertson
Miami-Dade Community
College

D. Wayne Silby
Groupware Systems

P. Michael Timpane
Teachers College
Columbia University

President
Russell Edgerton

LEGISLATIVE ACTION & ASSESSMENT: REASON & REALITY

COMPILED BY:
KATHLEEN MCGUINNESS

Proceedings from the Conference on
Legislative Action and Assessment:
Reason and Reality

Held in Arlington, Virginia

July 7-9, 1986

Hyatt Arlington, Arlington, Virginia

Sponsored by George Mason University
and
The American Association of State Colleges and Schools

Foreword

AASCU was pleased to co-sponsor with George Mason University an important conference on assessment, a subject that has long been of interest to our institutions. Questions of assessment have recently intensified, as commissions point out the weaknesses in our system of education, and the debate about access and quality becomes a critical focus of attention.

More than 15 years ago AASCU presidents issued a policy statement on "Quality and Effectiveness in Undergraduate Higher Education" which anticipated many of the current issues of assessment. The statement examined the serious need for a redefinition of educational quality at the undergraduate level and proposed measures of student achievement and institutional vitality. In his recent book, Achieving Educational Excellence, Alexander Astin cites AASCU as one of the most consistent supporters of the student development view of quality and quotes that 1971 statement:

> The customary measures of quality in most colleges and universities fail to deal with the impact of the institution on its students. However, if it is assumed that the teaching-learning function is the heart of the educational process, then the impact of the institution upon its students becomes critically important.

Astin also acknowledges AASCU's early concern with the question of multiple measures of effectiveness. He quotes from a 1976 statement, "Value-Centered Education and Moral Commitment":

> Colleges and universities have a responsibility to educate the whole student, not just serve vocational and scholarly interests. This becomes more evident as the complexities of the postindustrial society increase and as the influences of international events draw the world into a global community. Decisions, judgments, and perceptions are not as clear as they once were; and the consequences of action are more far-reaching. To function at the highest level possible, students will require knowledge, insight. and a foundation of values which have withstood the tests of history.

More recently AASCU has published, in conjunction with Northeast Missouri State University, a study called In Pursuit of Degrees with Integrity: A Value Added Approach to Undergraduate Assessment. That book describes a particularly effective university-wide program of assessment developed over a period of more than ten years to further the goals of quality for the individual student and for the institution.

Those goals were emphasized recently in testimony at the hearings sponsored by the National Governors Association Task Force on College

Quality. In that testimony Hoke Smith, chair-elect of the AASCU Board of Directors, defined assessment as "the measurement of an existing condition against a goal." He reminded the panel that the many goals in higher education and the many purposes of assessment make clear the complex problem of measuring performance. President Smith recommended that institutions make more positive uses of assessment and that individual systems be developed by the institutions.

In an interesting experiment a few years ago, ten AASCU institutions examined their missions and goals, then attempted to determine the effectiveness of their undergraduate programs. This Academic Program Evaluation Project, or APEP, shows that processes and results can be as diverse as the institutions themselves, yet accomplish the same purposes in assessing quality. The results of that experiment have just been published in a report, Defining and Assessing Baccalaureate Skills.

As these institutions have learned, it is not easy to establish valid and operational methods to determine educational quality. Yet we must recognize and meet the challenge if we are to achieve the potential of higher education for the future.

Allan W. Ostar
President
American Association of State
Colleges and Universities

INTRODUCTION

The idea for the AASCU/George Mason University conference on legislative action and assessment came about following a discussion with Dr. Ted Marchese, Vice President of AAHE, in which he expressed a desire for someone in the Washington area to host a conference which would move participants beyond "Assessment 101." The gratifying response of the participants leads me to believe that our conference accomplished this goal. I am grateful to Ted for his invaluable advice and assistance and to many other people as well, including Dr. George W. Johnson, President, George Mason University, and Dr. Allan W. Ostar, President, AASCU, for their support of the conference.

Several other college presidents advised me extensively. I am grateful to Dr. Harry Downs, President, Clayton State College; Dr. Charles McClain, President, Northeast Missouri State University; Dr. Jack Reese, Chancellor, University of Tennessee at Knoxville; and Dr. Hoke Smith, President, Towson State University. I want to extend special thanks to Dr. Evelyn Hively, Director, Academic Affairs Resource Center, AASCU, who coordinated the APEP panel; Ms. Kathleen McGuinness, Acting Director, Individualized Study Degree Programs, George Mason University, who assisted with the conference logistics and compiled the proceedings; and to my secretary, Louise Levinson, whose many contributions included worrying with me about whether or not there would be a conference.

The papers in this volume reflect the research and thinking of a number of educators who are concerned about the quality of undergraduate education and who are involved in the current assessment movement. The Division of Continuing Education at George Mason University is honored to have provided a forum for a provocative exchange of ideas on this significant topic.

Sally J. Reithlingshoefer

Sally J. Reithlingshoefer
Assistant Director
Division of Continuing Education
George Mason University

TABLE OF CONTENTS

Foreword....iii

Introduction....v

Defining and Assessing Baccalaureate Skills: Ten Case Studies American Association of State Colleges and Universities....1

Assessing Outcomes - The Real Value Added is in the Process
Trudy W. Banta and Homer S. Fisher....81

Assessment and the Reification of Learning
Clark Bouton....92

The Relationship Between High School Writing Experiences and University Admission Requirements
John E. Creeden....98

Academic Assessment: Implications for Faculty Evaluation
Robert DeBard, Ed. D. 107

Clayton State College
James Doig, Linda Greer and Joyce Swofford....114

Performance Funding for Tennessee Higher Education: How Does It Work and Where Did It Come From
Richard G. Dumont....121

Stimulating Local Assessment Initiatives by State Policy: Miami University's Response to Ohio Program Excellence
Curtis W. Ellison....134

The Assessment of Academic Quality: A Review of Existing Techniques with Emphasis on Their Validity and Feasibility in Public Higher Education
Willard F. Enteman and Pamela Irving Jackson....144

Where is the Value in "Value-Added" Testing?
Howard T. Everson....157

An Approach to "In-House" Test Development: For Major Field Assessment
Anthony J. Golden....167

Assessment in Undergraduate Education and Some Implications for Minorities
Freeman A. Hrabowski, Ph.D. 177

Miami-Dade Community College
J. Terence Kelly and John Losak....185

Evidence of Quality Assessment and Control in State Budget Documents
Jane Adams Lamb..191

Committee on Student Assessment
Frank Luth, Virginia Livingston, Andrew Kohen, Donald Corbin and Allen Lyndrup................................202

Creating Usable Program Evaluation
James H. McBath..218

State Control of Higher Education in Michigan: A New Scenario
T. Harry McKinney..225

An Overview of Institutional Evaluation
Richard I. Miller..231

The Role of Professional Judgment in the Use of Assessment Data
E. Thomas Moran..257

The Promise of Assessment: Linking Indicators to Outcomes
Thomas E. R. Redmon..263

Evaluation for Excellence: The Tennessee Plan
Linda Rudolph and James Nixon......................................273

Teaching Thinking: Implications for Institutional and Public Policy
Mark A. Schlessinger...280

The Uses of Assessment for Decision Making: A Primer About Northeast Missouri State University's Value-Added Program
Terry B. Smith...290

Evaluating Academic Programs
Hugh L. Thompson...309

Defining and Assessing Baccalaureate Skills: Ten Case Studies

A Report on the Academic Program Evaluation Program
of the
American Association of State Colleges and Universities
with major support from the
Fund for the Improvement of Postsecondary Education

Contents

Preface/

Introduction/

Chapter 1: Ball State University/

Chapter 2: California State University, Chico/

Chapter 3: North Adams State College/

Chapter 4: Ramapo College of New Jersey/

Chapter 5: Southern Illinois University at Edwardsville/

Chapter 6: State University College at Potsdam/

Chapter 7: University of Nebraska at Omaha/

Chapter 8: Wayne State College/

Chapter 9: Western Carolina University/

Chapter 10: Western Kentucky University/

Preface

Few of us in higher education would dispute the notion that college graduates should be able to demonstrate readily that they possess certain skills. Equally acceptable is the notion that among those skills must be the ability to communicate effectively, to engage in analytical and synthetical reasoning, to compute accurately and to solve quantitative problems of moderate complexity, and perhaps to develop a personal value system. Consensus begins to disappear, however, on definitions of each of these skills. In addition, if one further asks the classical who?, what?, how?, where?, and when?, the answers are not only elusive, but widely varying. The reports presented here describe the efforts of ten campuses to find answers.

These efforts arose from a project largely underwritten by a grant from the Fund for the Improvement of Postsecondary Education. Eventually known as the Academic Program Evaluation Project (APEP), the project was based on the assumption that a modular approach could in fact facilitate defining and answering the questions about academic skills. The project began with ten campuses serving as experimental sites at which a multistep model was to be tested. Under the direction of Marina Buhler-Miko, then the director of AASCU's Resource Center for Planned Change, teams of individuals from the participating institutions began to develop campus-specific approaches toward the ultimate objectives. There was a strong consensus among the project participants that every campus had unique circumstances calling for individualized approaches.

At the beginning of the project, the multistep process was available in limited detail. Later, it became clear that much more detail and an expanded theoretical base for each step were needed. Those needs should become apparent as one reads the reports from the ten campuses. Work in this regard has been proceeding through efforts of Robert Stakenas and Gary Peterson of Florida State University in collaboration with Ms. Buhler-Miko, who is now president of the Higher Education Strategic Planning Institute.

By disseminating the experiences of the ten APEP campuses, AASCU hopes to foster awareness of the necessity to give serious attention to the processes by which collegiate generic skills are to be developed; of the wide diversity of points of view that prevail on a single campus; of the need for efforts directed toward reaching consensus on what each campus means by each skill and its satisfactory development; finally, of the view that rigid national standards are next to impossible given the variety of institutions in this country. At the same time, every college and university must have a plan and program that will lead to demonstrable effectiveness in developing student skills as understood by each institution.

Harold Delaney
Executive Vice President
American Association of State
Colleges and Universities

Introduction

This compendium of case studies describes the efforts in the early 1980s of ten members of the American Association of State Colleges and Universities (AASCU) to evaluate the general education component of their baccalaureate degree programs by means of a uniform assessment formula.

The participants in AASCU's Academic Program Evaluation Project (APEP) were ahead of their time. Since their institutional evaluations took place, U.S. higher education has come under sharp attack, both from within and outside the profession, for allegedly fashioning curricula that are too utilitarian, failing to involve either students or faculty members systematically in "the learning process," and neglecting or ignoring the liberal arts. General education has suffered, the critics contend. Specialization dominates the undergraduate program, with lack of an institutionwide "sense of mission" the tragic result.

Clearly, the latter half of the 1980s will be a most difficult period for American colleges and universities. In response to the critics, many institutions will undergo some form of self-analysis. For a few, the experience--unhappily--may be too late to be helpful. For the more fortunate, it may occur--as it did for the colleges and universities involved in APEP--at a time conducive to campuswide change, resulting in enhancement of the learning process for both students and faculty members enrichment of the curriculum, strengthening of general education, renewed sense of institutional mission.

In one of the critiques of undergraduate programs published in late 1984, the Study Group on the Conditions of Excellence in American Higher Education, sponsored by the National Institute of Education, reported that "the quality of undergraduate education could be significantly improved if America's colleges and universities would apply existing knowledge about three critical conditions of excellence--(1) student involvement, (2) high expectations, and (3) assessment and feedback."

The study group was voicing precisely the concerns that motivated AASCU, in 1979, to initiate the Academic Program Evaluation Project (APEP), with partial funding by the Fund for the Improvement of Postsecondary Education (FIPSE). Many AASCU members perceived a need for evaluation of their baccalaureate degree programs, and APEP, as conceived, affirmed that there should be certain generic, recognizable outcomes of every such program. The project's objective was to establish a process by which all constituents of a college or university might specify the intellectual skills and abilities expected to accompany the earning of a bachelor's degree, prescribed in a manner acceptable to the constituents and in accordance with the aims of the particular institution. The process would necessarily include techniques by which the outcomes could be evaluated.

APEP's focus was to be broad rather than narrow. As an outgrowth of discussions on the meaning of the baccalaureate, the assessment of the quality of intellectual skills required for a bachelor's degree, and the reform of general education, the project was designed to provide means for evaluation of an institution's educational results. The descriptive term generic became useful for highlighting those intellectual skills and abilities which would be instructed, promoted, and required in all the disciplines, and which therefore might be said to characterize the recipient of a bachelor's degree, regardless of major or type.

Recognizing the need for a variety of plans from which other institutions could adopt those features applicable to their own circumstances, ten institutions were awarded FIPSE grants to implement the project on their respective campuses. These were Ball State University (Indiana); California State University--Chico; North Adams State College (Massachusetts); State University College of Arts and Science--Potsdam (New York); Ramapo College of New Jersey; Southern Illinois University at Edwardsville; University of Nebraska at Omaha; Wayne State College (Nebraska); Western Carolina University (North Carolina); and Western Kentucky University.

Under guidance of AASCU's Resource Center for Planned Change, the ten participating institutions formed committees of administrators and faculty members. While each institution would conduct its work as if the project were unique to its own campus, the committees formed a network for information and experience exchanges in conjunction with consultants and the staff of the Center. The first consultants, J. R. Warren of the Educational Testing Service (ETS) and Robert Stakenas of Florida State University, were joined later by Eldon Park of ETS, Gary Peterson of

Florida State, and Gary Woditsch of Bowling Green University (Ohio).

The Project: Outcomes of the Baccalaureate

APEP was based on the assumption that the outcomes of baccalaureate education can be expressed as generic intellectual skills; that these skills are the common denominators of the undergraduate curriculum; that they can be defined and broken down into measurable components; and that the extent to which they have been acquired by students can be evaluated. It followed, then, that if generic skills could be defined sufficiently to be measured, they could be used to evaluate the effectiveness of academic programs and, thus, of educational institutions. For the purposes of the project, a *generic skill* was defined as a complex set of intellectual abilities that can be generalized and are applicable to all the academic disciplines, and, ideally, would therefore be demonstrated by all graduating students.

The idea of evaluating undergraduate education in relation to transinstitutional, generic outcomes developed over many decades, but special attention was focused on actually implementing such an evaluation when, in 1978, the American Council on Education's Task Force on Educational Credit and Credentials reported the following recommendations:

> Associate degrees and Bachelor's degrees should attest to at least the following three types of accomplishment:
>
> **1.** Accomplishment specified by the awarding institution as necessary for the development of a broadly educated person, including familiarity with general areas of knowledge;
>
> **2.** Competence in analytical, communication, quantitative, and synthesizing skills;
>
> **3.** Accomplishment in a specialized area of study covering a set of integrated learnings requiring analysis, understanding or principals that have judgmental application, and a theoretical knowledge base.
>
> Each credential-granting institution should clearly define, to the extent possible, the meaning of the certificates and degrees it awards.

Following upon these recommendations, J. R. Warren of ETS undertook research to discover whether it was possible to measure such "generic" outcomes as analysis, communication, quantification, and synthesis. At the same time, the Center Associates of AASCU's Resource Center for Planned Change agreed that the same four generic skills should become the common base, yielding individualized interpretations, for the ten APEP institutions. A fifth skill--valuing--was also identified and could be included at the option of each participating institution.

It should be noted that APEP was not the only project expressing educational outcomes as generic skills. Research reports in educational psychology had suggested that intellectual skills traverses subjects and disciplines, that learning of skills proceeds in a hierarchical fashion, and, thus, that individuals have varying levels of skill development that are independent of subject-matter mastery. These fundamental developments in learning theory brought on the "interdisciplinary" and then the "competency-based" curricular trends of the 1970s. During this time, the College Board began its ten-year Equality Project to address preparation for college by defining the learning skills students should have in order to succeed in college. School and college faculty members and other educators natiowide participated in identifying six basic academic competencies: reading, writing, speaking and listening, mathematics, reasoning, studying. In a second phase of the project, six basic academic subjects were identified for secondary education: English, the arts, mathematics, science, social studies, foreign language. More recently (1984), AASCU published *In Pursuit of Degrees with Integrity*, describing "a value-added approach to undergraduate assessment" adopted by Northeast Missouri State University.

By using generic skills to define the fundamental outcomes of baccalaureate education, it was possible to design an evaluation process that would adjust for different institutional goals and involve faculty members from all disciplines. The five-stage evaluation process used by the APEP participating institutions is outlined in the next section of the Introduction. Once the ten institutions had identified the generic intellectual skills their respective general education programs sought to produce in their students, the evaluation process consisted of defining each skill in measurable terms, selecting instruments and procedures for measurement, assessing student acquisition of generic skills, and evaluating the results in relation to the academic program. Policy implications emerged as the evaluation proceeded at each institution.

Because the APEP institutions had agreed on communication, quantification, analysis, synthesis, and valuing as the relevant generic skills for the project, at each campus the work began with attempts to define these skills in specific, measurable terms. The campus project directors and the interdisciplinary teams or committees proceeded to construct "conceptual definition, component subskills, observable behaviors and levels of of performance" for each generic skill. They were assisted in this task by the project consultants. Although AASCU's Resource Center staff provided guidelines for the process of defining generic skills, each campus approached the task in its own fashion, as the ten case studies prepared by the campus project directors reflect. Similarly, each institution followed its own procedures on other phases of the project, from the composition of committees to the method of decision making.

The Five Stages of APEP: The Evaluation Process

The five stages of the Academic Program Evaluation Project were designed to direct the attention of faculty members and administrators to key issues in the assessment of generic skills. The stages were intended to provide guidance, not a prescription, and the scope was deliberately planned to be broad enough to accommodate diverse institutional missions, curricula, instructional styles, standards, educational policies, planning, and measurements.

Faculty and administrative representatives of institutions participating in the project were expected to work together to

o identify institutional aims in terms of generic intellectual skills

o define each of the generic skills valued by the institution in measurable terms

o determine where and how well those skills were being taught

o pinpoint curricular and instructional instances in which generic skill performance was below required standards

o recommend policy and procedural changes designed to raise performance levels in any areas of low generic skills.

Procedurally, these basic tasks were conceived as a developmental process, and thus developmental terms seemed most appropriate in suggesting the steps through which the participating institutions would pass. The APEP process is, therefore, most easily described in terms of "stages," with certain tasks to be accomplished during each stage.

Stage I: Developing Definitions of Generic Skills

An institution that seeks to evaluate its academic program typically begins by defining the generic skills it seeks to encourage in its students. Although APEP focuses on the specific skills of analysis, synthesis, quantification, communication and valuing, the skills chosen could be any generic skills valued by the faculty. Formulated in concise, conceptual terms, these skill definitions are the guides for all subsequent assessment steps.

By definition, generic intellectual skills are cross-disciplinary and were considered, at least initially, in discipline-free terms. They reflect the essential and common characteristics of thinking processes that occur in every field of study. Once the abstract statements of capabilities are agreed on, the work then proceeds to the identification of subcomponents or subskills that can demonstrate the capabilities. After these definitions are submitted for faculty consideration and approval, Stage II begins.

Component Tasks

1. Determine the purpose of an evaluation project, whether as broad as the entire baccalaureate level or as specific as selected professional programs. Establish the intended audience for project results. Determine the project scope in order to identify the period of time (e.g., freshman and sophomore years), groups of students and programs, and any special attributes.

2. Review the institutional mission to identify the generic skills inherent in the stated purposes of the college or university.

3. Prepare conceptual definitions of the generic skills and subskills that are measurable. Gradations in performance levels need to be described. Distribute these definitions and associated features to the faculty for critical response and ultimate acceptance.

4. From among the defined generic skills and subskills, select those for inclusion in the evaluation. (Ordinarily, it would be too ambitious and predictably unwise to pursue measurement of all the skills.)

Stage II: Identifying Performance Indicators, Criteria, and Testing Procedures

The definitions of generic skills become useful only if observable indicators are discovered. Such indicators should be sought in the intellectual behaviors demanded of students, indicating various levels of competence, in various desciplines. Generic skills cannot be observed directly, and these "indicator" behaviors or performances permit evaluators to draw inferences about the skills.

The major task becomes one of identifying and developing testing procedures to evoke student performances indicative of generic skill development. Statements describing levels of performance--medium, high, low--must be formed and criteria established to judge the level of student development. Because generic skills are multifaceted, it is unlikely that they could be adequately assessed by a single procedure. Multiple performance measures with varied complexity and subject-matter content of the assignment present students with opportunities to demonstrate their competencies in differing intellectual situations.

Component Tasks

1. On the basis of the project mission, select and validate the scope, skills, and content of the project. "Scope" denotes the context in which the skills are developed (e.g., a course, a major, a baccalaureate degree, life in general). "Skill" is the generic skill to be assessed. "Content" refers to the complexity and breadth of the domain wherein the skill development is believed to occur (e.g., a topic, a discipline, a cluster of fields, an entire program).

2. Design the assessment procedures. Direct measures of performance include all forms of written, oral, and artistic works whereby the student's product or his contribution might be assessed. Indirect measures can be used to evaluate a student's choice of responses that are indicative of future performance.

Beyond selecting the assessment procedures, take steps to ensure the availability of materials, resources, and guidance necessary to prevent irrelevant handicaps to student performances.

3. Develop specifications for tests and test items. The preparation of locally developed tests calls for available professional talent to guide and validate the product. This effort requires great support from the faculty. Pilot test all instruments purchased or devised.

4. Establish a general policy and process for all tests, whether home developed or purchased commercially. The process should provide for a general analysis of the merits and liabilities of locally produced *vs.* commercially available tests.

Stage III: Assessment of Students and Programs

There are two tasks for this stage: assessing of generic skill development among students, and determining components of an academic program in which a student has the opportunity to acquire or improve generic skills.

Because it is not feasible to measure every student's development and not possible to observe every learning experience, decisions are needed on which students to sample, how many students to sample, and the extent of academic programs experienced?

A plan for data collection, analysis, and interpretation must be drawn up.

Component Tasks

1. Formuulate the learning outcome evaluation design--i.e., the framework for decisions about where, when, and from whom data will be collected. Recognize that a control group and rigorous sampling procedures would not be possible initially.

2. Specify the instructional events that develop generic skills; that is, in what parts of the instructional program generic skills likely are to be developed.

3. Identify sequences of courses or other identifiable learning experiences and endeavor to learn which of the skills is likely to be affected by the elements of sequences. Once the identity is established, a sampling procedure can be found and data collection begun.

Stage IV: Evaluation--Judging Student and Program Performance

The data accumulated in the assessment of student progress must be interpreted against standards presumed relevant to the institutional mission and student body. Traditionally, the standards are those fixed by the faculty.

Student performance standards can be relative--i.e., compared with students in other places or with students of a prior year. Standards can also be absolute, as defined in faculty convictions about the goals of a baccalaureate.

Comparing performance data with the standards establishes whether the standards have been met; whether the intents and the observations correspond; and whether the gaps discerned in student performance are serious enough to warrant action.

Component Tasks

1. Arrange all data accumulated relative to programs and students for analysis. It is often convenient to divide the information into antecedents (givens), transactions, and outcomes in order to analyze and describe of what was intended and what was observed.

2. Determine the congruence between what was intended and that which was observed. Lack of congruence could indicate that the learning assignments did not effect the desired skills, or that the applied measurements were not adequate to assess the skill in question.

3. Determine the logical and empirical contingencies. The former are based on prior experience and show reason in the planning of instruction and assignments and are conguent with the course objectives; tests at the end should also be congruent. Empirical contingencies require a kind of audit of faculty members and students done independently to judge the performance of both--ultimately, to discover any lack of congruence in skill development.

4. Compile a list of discrepancies and performance gaps.

5. Determine standards for making judgments. A comprehensiveprogramassessmentwouldjudgemorethan student outcomes; it would focus as well on the "givens" and the entire intellectual strategy. Hence the standards applied to the whole, including the levels of expectation of students, becomes the heart of the assessment process. Setting standards is an act of judgment, and participants should exercise great care in order to ensure fairness and responsibility.

6. On the basis of standards, make a judgment of worth--that is, a comparison of what was observed and the prescribed standards, with discrepancies noted and reviewed for levels of importance. Those deemed critical are reserved for consideration in Stage V.

Stage V: Policy Analysis and Decision Making

If discrepancies were noted in Stage IV, determine the causes. Did the discrepancies arise because of limited resources or inappropriate use of resources? Could the cause be traced to admissions, curricula, instructional practices, or some other variable? If the causes are found, remedies can be devised and assessed for potential effectiveness and feasibility.

In this stage, consider alternatives for the remedy or the neglect of discrepancies. A decision to apply a remedy could cause the institution to revisit another stage. For example, if a decision is made to redraw original definitions, Stage I would be reentered. If different testing procedures is indicated, Stage II would be reentered. If the definitions and testing procedures remain the same, then learning outcomes and program effectiveness could be reviewed in Stage III.

Component Tasks

1. Determine the reasons for performance gaps in observed outcomes. In some instances, the component steps outlined in Stage IV could assist in this determination.

2. Compare the cost of removing the gaps with the consequences of ignoring the gaps.

3. Arrange the gaps in learning outcomes according to importance and select gaps on which actions are to be taken.

4. Determine policy changes required for removal of the gaps. Revised goal statements relative to the generic skills are often necessary. Sometimes academic goals require revision to clarify expectations about the skills.

5. Make changes in instructional practices necessary to remove gaps. Significant changes in student and faculty role behavior are frequently needed.

6. Encourage participation of the faculty in making decisions about policy and instructional changes. This step has proven crucial to success.

7. Devise a time schedule for completion in order to facilitate evaluation of the effectiveness of the changes.

Following are descriptions of the APEP experiences of ten AASCU institutions, together with a summary of the policy and administrative issues raised by academic program evaluation on the participating campuses.

Ball State University (IN)

Institutional Profile

Ball State University is a comprehensive, publicly assisted institution of higher education that aspires to excellence in teaching, research and other scholarly productivity, and public service. The university's missions and constituencies are regional, statewide, and national in character, depending on the program or activity. The university also engages in a selected number of programs distinctly international by virtue of their location and clientele.

Ball State University is an academic community substantially residential in character but nonetheless one in which outreach activities, continuing education, and the provision of education in a variety of off-campus locations loom large. Moreover, the university is dedicated to the pursuit of truth and the provision of a liberal education to all its students regardless of their majors or locations.

The university offers a comprehensive range of academic programs at the associate, baccalaureate, and master's levels, as well as doctoral programs in areas in which it has special competence and the State of Indiana has special needs. In addition to the core academic programs in the arts, sciences, and humanities that every comprehensive university provides, Ball State also offers a broad range of professional programs that not only prepare students for specific occupational and vocational interests but also provide them with a liberal education. Important among these professional offerings is a complete and diverse range of teacher education programs that build upon the university's traditional strengths. These professional offerings also include a widening circle of programs in such areas as actuarial science, allied health sciences, architecture, landscape architecture, urban and regional planning, nursing, bioenergetics, a wide range of business and management-related disciplines, computer science, library science, public administration, social work, and speech pathology. Ball State University is committed to furthering professionally oriented, programmatic growth in those areas wherein it exhibits competence and there is demonstrated need.

The emphasis of the university's major academic programs during the next decade will be concentrated in areas that support the provision of human services; the preparation of liberally educated students in professional and vocational fields; and the solution of societal problems, including those pertaining to energy, high technology, resource management, health, and the quality of life. Some of these programs will be regionally oriented by virtue of the university's special relationship with its geographic area. Others, such as architecture and planning, teacher education, and the bioenergetics graduate program, will reflect not only state needs and constituencies but national ones as well.

The university's policies governing admission and retention of students are based on a preference for students who exhibit either immediate ability or strong potential to benefit from higher education at Ball State. Hence, the university is selective and does not accept for admission all students who wish to enter. In certain areas, the university's admissions policy is highly selective. The university's interest in students of high ability and potential transcends variables such as age, handicap, occupation, sex, race, location, and part-time status. Exceptionally well qualified undergraduate students may enroll in the university's Honors College. The Academic Opportunity Program provides special academic assistance to a finite number of less-qualified but promising students. Pursuit of individual goals by disadvantaged students is encouraged and facilitated. All of these opportunities ultimately reflect and help maintain the standards of excellence of the entire university.

The continuing maturation of Ball State University will lead to a broader definition and assignment of responsibilities for both the faculty and the university itself. While quality teaching will remain the most highly valued focus for the university's faculty, increased attention and resources will be given to research and other scholarly productivity and public service activities. The university will obtain increasing proportions of its financial support from external sources, both public and private. The result will be a balanced and complementary relationship among teaching, research and other scholarly productivity, and public service that will distinguish Ball State from both land-grant research institutions and liberal arts colleges, to which research and public service are less important.

The mission of the university will undergo continuous scrutiny. Significant changes will become necessary as new students, fields, programs, techniques, interests, and technologies affect the institution. Yet the future will not alter the university's firm commitment to the highest possible quality in its activities of teaching, research and other scholarly productivity, and public service, or to the preservation of a positive and attractive environment for learning and living.

Background

When the American Association of State Colleges and Universities announced its intent to assist ten member institutions in evaluating program effectiveness in general education, the offer was seen as a timely opportunity for "fulfilling a deferred commitment." Although the mood among institutions of higher education generally was to move away from the "loose" programs of the preceding decade to some kind of core curriculum, patterns of change reflected little direct relationship to the development of generic skills assumed to be the outcomes of particular programs and, thus, the hallmarks of an educated person. With its special focus on the intellectual tools of analysis, synthesis, communication, quantification, and possibly valuing, and with the promise of special assistance in assessing these skills, the AASCU invitation sparked notable interest.

APEP, with its primary emphasis on assessing skills rather than particular subject achievement, carried a special appeal to those on our faculty who felt that all undergraduate instruction should reflect a sensitivity to developing the generic skills of not only communication and quantification but analysis and synthesis as well. Such a view was reinforced by the statement of purpose for general education in the Ball State University catalog, which asserts that the aim of general education program is "to provide undergraduates with the concepts, understandings, skills and values necessary for educated men and women to live purposefully in modern society." Moreover, the statement asserts that "the distinct courses in general education share common goals in the sense that they all should promote the spirit of inquiry, relate knowledge from various fields, and encourage continued liberal education on the part of the student."

With the above institutional philosophy in mind, it was thought that Ball State University's participation in APEP might provide evidence of success in terms of one of the institution's major goals--namely, to provide general education. The nature of the skills to be measured suggested further that the institution might determine whether its program promoted a spirit of inquiry and related knowledge from various fields. The project's concern for assessing such outcomes as the ability to analyze and synthesize was particularly attractive to the many faculty members who recognized the importance of these outcomes but who lacked sophisticated means for their assessment.

Apart from these direct outcomes related to the goals of general education, the university viewed the project as an opportunity to find answers to certain policy questions common to its total undergraduate curriculum. First there were the policy questions of the meaningfulness of separating general education from special-interest education, i.e., of answering the query, To what extent are sharp distinctions warranted between general and specialized programs? It was thought that participation in APEP project would reveal the differences in educational outcomes between those students pursuing highly structured professional progams and those whose major concentrations allowed greater freedom of selection from within the general studies offerings.

A second policy question of timely concern was that of balance between constraint and variety; in other words, to what extent was the program of general studies ensuring the realization of total institutional goals amid increased pressure for greater autonomy for various schools and colleges within the university? At Ball State the general education of students varies widely in relation to use of electives recommended by their respective major departments; therefore, the university was faced with legitimate policy issues concerning centralization and decentralization of curricular and instructional practices. The program of general education, in becoming the focus of these tensions, offered a clear case of policy delineation. One important outcome of the current AASCU-APEP project would be to ascertain the optimal balance of constraint and variety in a program that would also meet standards of effectiveness in the four areas of concern in this evaluation.

Establishing the focus of responsibility for certain educational outcomes was the third policy matter that might be settled through participation in the consortium of AASCU-APEP institutions. To what extent and by what means could the institution place the responsibility for developing communication and quantification skills in all academic units rather than exclusively in the departments of English and mathematics? Institutions with open admissions standards are receiving an increasing proportion of students lacking basic skills of computation and communication. Because current responses to this dilemma have proven inadequate, Ball State University recognizes APEP project as an opportunity to involve a wider segment of the faculty in addressing the problem.

Implementation

Defining Generic Skills

Shortly after it was announced in October 1979 that Ball State University had been chosen to participate in APEP, then Provost Richard W. Burkhardt extended an open invitation to the faculty to attend a general session explaining the intent and goals of APEP. Many faculty members responded to the invitation and were subsequently invited to attend a general organizational meeting.

At this meeting, attended by some forty faculty members, four groups were formed, each charged with developing conceptual and operational definitions for one of four skill areas under study. These four groups continued to meet both individually and in plenary sessions throughout the winter quarter of the 1979-80 academic year, refining and reworking the definitions. Their efforts culminated in a workshop held at the University Kitselman Center in February 1980. During the workshop, much discussion focused on defining the levels of performance for each of the four abilities and on planning an approach to identifying when and how these abilities were being taught in the general education curriculum. Throughout March and April of that year, faculty members continued to meet in general sessions, improving not only on the conceptual but also the operational definitions. Six basic assumptions informed and guided the thinking of the four committees:

- that the four thinking skills defined were but a few of a wider array of important skills and that definitions can be developed only in such a context (for example, analysis and synthesis are only two of six for the Bloom taxonomy)
- that thinking skills can be defined in behavioral terms and observed, described, and measured in some way
- that given the present constraints of the state of the art of educational measurement, the definitions, in order to be useful in this project, need to be susceptible to "pencil and paper" measurement
- that the context adopted by the Ball State University faculty for defining these skills is primarily cognitive rather than affective
- that such skills can be defined more distinctly at basic levels and that definitions become interrelated at advanced levels
- that such skills present infinite possibilities for development throughout life.

By the end of April 1980, each of the four committees had settled on the following basic conceptual definitions.

Communication

The committee on communication defined this ability as receiving and sending messages through such processes as reading, listening, writing, and speaking. The committee pointed out that although nonverbal communication was not included, its importance was not to be denied. Furthermore, it was noted that the component skill of effective speaking was perhaps the skill least susceptible to the type of assessment associated with the APEP Project.

Analysis

The committee on analysis developed the following statement: analysis is the process of separating a communication into its constituent parts to make clear the order of and relation among ideas expressed; to indicate the organization of the communication and the way communication manages to convey its meaning; to show the bases for and arrangement of the communication, both explicit and implicit. One issue that surfaced in constructing the definition of analysis involved the notion that there is a context, not developed in the definition presented above, in which the skills of analysis and synthesis are complementary components of the broader skill of "problem solving." A question that also promoted significant discussion was that of whether it was possible to widen the scope of the skill of analysis beyond a "communication" and still render the skill susceptible to measurement in the traditional sense?

Synthesis

The committee dealing with synthesis presented its conceptual definition by identifying six operational components:

- to organize unrelated facts and ideas into a framework
- to introduce new ideas into the perspective of an existing framework
- to speculate on new frameworks within existing systems
- to extrapolate from known to unknown situations
- to recognize connections between theory and practical applications
- to integrate new facts and ideas with personal experience.

Two unresolved questions emerged from the deliberations concerning the skill of synthesis. One was whether the parameters of APEP reduced the notion of synthesis to verbal communications. A second was that of whether the skill of synthesis should be defined in a way that complemented the definition of analysis.

Quantification

The quantification committee, after much discussion, accepted the following definition: the ability to understand and use symbols and diagrams to express quantity and relationships. In its discussions on quantification, the committee wondered whether the definition finally determined was too closely associated with the discipline of mathematics. Additional musing led to the suspicion that advanced quantitative skills could well be very similar to the skills of analysis and synthesis.

Assessment of Skills

Upon acceptance of the operational definition of skills by the participating faculty members, a committee was selected from the group and charged with the task of constructing a research design for the project. A preliminary evaluation design emerged in spring 1980. A pilot program was proposed to test the validity and applicability of test instruments sent to the university by the Educational Testing Service (ETS) under the direction of Jonathan Warren. With a final evaluation design for the project approved by the committee in early October 1980, the assessment of skills was undertaken.

Purposes

As the sequential stages of the APEP paradigm unfolded, the general purpose of the project was increasingly refined. The unique nature of the general education program at Ball State University together with the time constraints of the APEP project caused the Ball State University planning group to view the initial implementation year of the project as a formative step in continuing evaluation of the general education program. The purpose of the first implementation year was articulated as follows:

> To assess entry-level and exit-level student performance in four generic skills (i.e., analytic, synthesizing, quantitative, and communication) to initiate a program of continuing evaluation of the effectiveness of the Ball State University general Education Program, as it functions in concert with majors in the major curricular divisions of the University (i.e., humanities, social and behavioral sciences, and sciences and mathematics, or business, technology, and applied fields).

To achieve this overarching purpose, Ball State made the following decisions on the nature, scope, and focus of the assessment mission:

- to focus the investigation on the four generic skills mentioned above and to omit the generic skill of valuing
- to use criterion-referenced (competency) measures to assess high- middle-, and low-levels of skill performance
- to adopt and validate available instruments that were relatively "discipline-free"
- to supplement the basic measures with various unobtrusive data such as entrance test results and achievement of various levels of academic honors
- to focus the first-year implementation study on outcome assessment of both freshmen and seniors
- to use a limited program portrayal approach for the assessment of skill growth of freshmen who would take various combinations of five or six general education courses during their freshman year and to study the difference between seniors who pursued a structured program of general education and those who did not do so.

It was anticipated that this approach would facilitate a number of significant outcomes: (1) a provision that would allow judgments about the entry and exit generic skills of current Ball State University students as well as a basis for making some limited judgments about the effectiveness of certain combinations of general education courses, (2) judgments about the validity and scope of the various measures used, and (3) a data base not currently available to facilitate future evaluation efforts. It was also hoped that this approach would stimulate interest in carrying out increasingly focused outcome and program portrayal evaluations in subsequent years.

Types of Skill Assessment

Standardized, multiple-choice instruments prepared, field tested, and revised by ETS were chosen to measure the four generic skills. The specific forms for communication, analysis, synthesis, and quantification were ETS Form 14 Revised, ETS Form 34 Revised, ETS Form 44 Revised, and ETS Form 27 Revised, respectively. These particular forms were chosen from three of four alternate forms for each skill by the research design committee, which hoped to reduce the total testing time to one hour. The committee did recognize that the brevity of these instruments might raise questions about their power, reliability, and validity.

Specific Cohorts Selected for Testing

Five cohorts of students were defined for testing in the assessment stage of the project: (1) 375 matriculating freshmen, (2) 91 of the 375 freshmen seven months after matriculation, (3) 260 randomly seleted seniors in their final quarter of studies, (4) 39 seniors who had been admitted in the "distinction" category and who had completed the structured Honors College curriculum of general education, and (5) 44 seniors who likewise had been admitted in the "distinction" category but who had not pursued the honors curriculum. With the exception of the matriculating freshmen, who were tested in the autumn quarter, all other groups were tested late in the spring quarter of 1980-81.

Because the major thrust of the project was descriptive and analytical, the following questions were viewed as generally more helpful to the evaluation than specifically formulated hypotheses:

- Are there distinct differences in skill performance in freshmen after a year of general education?
- Are there differences after a year between those freshmen who complete a substantial number of general education courses and those who do not?
- Are there distinct differences between matriculaters and graduating seniors?
- What proportion of the beginning freshmen have attained *low* skill performance, *medium* skill performance, and *high* skill performance in each of four areas?
- What proportions of the seniors have attained *low* skill performance, *medium* skill performance, and *high* skill performance?
- Are there significant differences in seniors of equal ability between those who pursue a structured general education and those who do not?

Study Results

Communication

On the test of communication skills the freshmen showed some gains between the autumn and spring quarter, but the differences were not significant at the 0.05 level of confidence. A formal analysis of scores made by freshmen who had completed eight or more general studies courses and those who had completed five or fewer such courses showed a slight difference between groups, but not significant at the 0.05 level of confidence. When beginning freshmen were compared with seniors, the latter group reflected a higher skill level. The differences were found to be significant at the 0.001 level of confidence. When arbitrary standards were set for low, medium, and high performance for the communication test, the range and distribution of scores for freshmen and seniors reflected the differences reported above. A comparison of differences between two groups of seniors--those completing the more structured general education program of the honors curriculum and an equally able group not pursuing that curriculum--showed the former group had higher scores than the latter. However, differences were not found to be significant at 0.05 level of confidence.

Analysis

On the test designed to measure the intellectual skill of analysis, the freshmen showed statistically significant gains between the autumn and spring testing. The limited sample (N = 28) of freshmen having taken eight or more general studies courses was compared with those taking only five or fewer (N = 29). The group having completed five or fewer general studies courses scored significantly lower than the group completing eight or more at the 0.001 level of confidence. On this test seniors performed better than freshmen, the difference being significant at the 0.001 level of confidence. Again, when arbitrary standards were set for various levels of performance, the gains of freshmen and the differences between freshmen and seniors were obvious and consistent. When the two groups of academically able seniors were compared, the average mean score of the Honors College group was markedly higher than that of their classmates of equal ability. Differences were found to be significant at the 0.001 level of confidence.

Synthesis

On the test designed to measure the skill of synthesis, gains made by freshmen between the autumn and spring were significant at the 0.001 level of confidence. The informal analysis of two groups of freshmen with varying experiences in general education showed the group with more such courses to have only slightly higher increments than its counterpart with fewer general courses. Seniors, however, showed significantly higher scores than did freshmen on this test. When the range and distribution of scores were arbitrarily grouped in reference to low, medium, and high standards of performance, the differences both between freshmen groups and between freshmen and seniors reflected the earlier findings. Again the Honors College seniors performed better than their classmates of equal ability. However, differences between groups were not found to be significant at the 0.05 level of confidence.

Quantification

Of the four skills measured in the study, quantification reflected the greatest difference within the freshmen cohort and between the freshmen and seniors. Differences were significant at the 0.001 level of confidence, and t-values were the highest of the four areas of comparison. Statistical comparisons showed the group of freshmen with five or fewer general studies courses to have a lower average mean score than that of the group with eight or more such courses. However, differences were not significant at the 0.05 level of confidence. Differences between average mean scores of the Honors College seniors and their counterparts were noticeable but not significant at the 0.05 level of confidence.

Conclusions

To find significant incremental differences between the freshmen tested in the autumn and those retested in the spring on three of the four tests is cause for general satisfaction with the experience provided students in the initial year of studies. Without national norms for the tests administered, though, it is difficult to know whether the increments are sufficient to make positive evaluations of that experience. Furthermore, the limited samples of students having varied contact with general courses precludes anything beyond conjecture that such courses are more likely to develop the skills of concern in this study. The finding that the differences of test scores on the communication test were not significant provides little help in assessing the value of freshman English composition courses. The rather limited nature of the test, which emphasizes reading comprehension to the neglect of other important aspects of communication, points out rather specifically the need for identifying and using a more adequate test for measuring this important skill.

In contrast, the test for quantifying skills was perceived to measure the skills specifically sought in the introductory courses in mathematics. The data related to performance on the quantification test, on which differences were perceived to be most significant, also present a quandary. A formal analysis of the groups with varying numbers of general courses suggests that the panel of general studies courses have little more impact than would a collection of unspecified electives.

As with the increments shown by freshmen, the favorable differences achieved by seniors over the freshmen suggest the positive impact of the undergraduate experience. Again, comparison with normative data would have provided greater assurance that the differences were sufficient to warrant a high level of satisfaction with the outcomes. Also, it is not known how much the difference might be attributed to normal maturation. The fact that on all four tests the increment shown in the freshman year is equal to the differences between freshmen in the spring and seniors may suggest a leveling off of the development of these skills as the student progresses through four years of undergraduate studies. This diminishing rate of change should at least be examined for its fuller meaning.

Probably one of the more provocative questions to be studied in interpreting the data from test scores deals with the differences found between the two cohorts of seniors who had been admitted four years earlier in "Distinction" category. A control group of forty-four seniors in this category had pursued a general studies program characterized as having little structure and consisting of electives in the various distribution areas of humanities, social sciences, and natural sciences. In contrast, the group of thirty-nine seniors who were completing the Honors College curriculum had pursued a more structured program of general education. This program comprises two values-oriented symposia in the first year, a three-course humanities sequence in the second year, two informal colloquia in the third year, and a capstone honors thesis in the last year. On all those tested, the honors group performed better than their classmates with comparable entrance test scores. Further analysis of these differences could provide the investigators with significant information on which to base assessments of this particular pattern of general studies.

Policy Development and Decision Making

Three fundamental policy questions confronted the university at the outset of this study: (1) To what extent are sharp distinctions warranted in separating general education from specialized studies? (2) How should a program of general studies be structured to achieve a balance between constraint and variety? and (3) Where should the university place the locus of responsibility for developing and appraising the program of general studies? Participation in APEP provided increased insight into these pervasive issues and generated a greater level of faculty consciousness of the need to make a series of recommendations in response to their new understanding of the comprehensive problem of improving general education. This section of the case study summarizes the nature and scope of this heightened consciousness and proposes those general and specific policy recommendations which seem in order.

Response to Basic Issues

The question of maintaining a distinct separation of general and specialized education forces educators to reconsider the overall purposes of undergraduate education. the unique claims made for common learnings, and the appropriate emphasis on vocational preparation. There was nothing in the experience of the Ball State University faculty task group to suggest that the aims of general education would be served automatically in an undergraduate curriculum increasingly characterized by specialized education. Instead, the discussions on the campus in small groups and plenary sessions supported not only the concept of a visible program of general studies but also one that would foster the development of those intellectual skills that were the particular focus of this study. In addition, there seemed to be full recognition of the need for making general education a university-wide enterprise of concern to all, even though primary responsibility for its delivery may rest with the faculty of particular disciplines. What was needed, it seemed, were better mechanisms for ensuring that what was provided in the name of general education was indeed all that it promised to be. Many of the recommendations offered by the group emanated from the conviction that the cause of general education should not be abandoned but instead reinforced consciously by the faculty and administration. This would require greater attention to the structure of the program, to quality review by various governance units, and to continuous evaluation by an administrative unit with resources sufficient for the task.

The issue of structure in an ideal general studies program pervaded the deliberations of the faculty throughout the period of this study. Whatever consensus achieved was limited to a general feeling that the existing program lacked sufficient structure to achieve its purposes. A majority of participants considered the surfeit of existing options so amorphous that a single program for all students was difficult to identify. When questions were raised about the parts of the curriculum that develop generic skills, it was apparent that, with the exception of the skills of communication taught in two required English courses, there was no component of common learning; on the contrary, each student's program of general education constituted a mixture of conveniently scheduled electives, directed electives in support of the majors, and whatever additional courses might be needed to fulfill distribution requirements. This limitation was particularly obvious at that point when the group wanted to compare the performance of several student cohorts with different patterns for meeting these requirements. There were as many patterns as there were students and, with the exception of the Honors College seniors, no consistent patterns prevailed.

Although no formal proposals on structure emerged from the APEP experience, there seemed to be support for a general education program characterized by a common core of specific courses complemented by electives from specified groups of courses in the broad fields. Evidence from the study of skills developed by seniors who had participated in a structured sequence of Honors College courses supported this inclination. As in most discussions of curriculum structure, there was an articulate minority who eschewed any required formal organization but instead brushed aside the issue of structure for an emphasis on the instructional strategies within courses.

Inasmuch as the university launched this study with some concern about the locus of responsibility for developing and evaluating the general studies program, it is appropriate at this point to assess the experience in this regard. Apart from the general argument that such studies, and particularly the development of generic skills, should be a responsibility shared widely in the university community, there was wide recognition of the need for regular and continuous assessment of program and student performance. A review of some of the major assessment problems will illuminate this need in greater detail.

Shortly after the APEP task force began to work through the various stages of the evaluation paradigm, it became apparent that the university did not possess in a single administrative unit all the resources needed for pursuing a project of this scope. Although responsibility for conducting the project rested nominally in the Office of Undergraduate Programs, the sources of expertise in measurement, design, and statistical analysis were located elsewhere within the university's organizational structure. Furthermore, an emphasis on voluntary participation, made primarily because the institution was undergoing rather frequent turnover in top administrative leadership, meant that the task force included for the most part faculty members and administrators who, though committed to the project, lacked certain necessary skills in evaluation. The experience of dealing with these deficiencies has led to the conclusion that the university needs to place clear responsibility for this type of assessment in a single administrative unit possessing the necessary staff and inclination to conduct such studies on a routine and continuing basis.

To establish an assessment center whose responsibilities would be to assess and certify student attainment of the generic skills, the university would need to reorganize existing units that share but do not claim full authority or responsibility for these matters. Presently the university maintains the following units, which collectively possess the necessary staff: Office of Institutional Studies, University Computer Center, research computer design unit of the Office of Research, Director of Analytical Studies, the psychometric staff of the Office of Counseling and Psychological Services, and the Director of Student Affairs Research. These offices, however, are administered by the president, provost, and the vice president for student affairs. Should the university decide against reorganizing its resources into a new unit, there would need to be an unmistakable mandate from the chief executive officer clearly identifying the responsible coordinating unit and the cooperation necessary from other support units.

The choice of instruments for use in this study represented a compromise between aspiration and expediency. Although the limitations of the scales developed by the ETS were readily recognized, the will to develop better ones within the available time frame was lacking. After examining the COMP materials produced by the American College Testing Program and those of the McBer Company, the faculty opted for the shorter scales under development by ETS. The choice of these instruments was predicated on the anticipation that national norms might be made available during the course of the project. Unless an institution has sufficient facilities and faculty interest to develop "home grown" measures that fully reflect their objectives related to generic skills, there is little choice except to use currently available commercial instruments. Many in the group saw in this dilemma an additional area of professional development that would be served by the university's Office of Instructional and Professional Development.

Although certain compromises were made in selecting the various samples of students to be tested, subsequent analysis suggests that fairly representative cohorts were included. The Research Design Committee would have preferred to test the entrance skills of matriculates during the orientation process in summer 1980 but had not progressed sufficiently toward a selection of measurement instruments to accomplish the task by that time. Consequently, it was necessary to choose as test sites those sections of English and mathematics courses with high enrollments. Similarly, the choice of courses with a high enrollment of seniors as sites for exit testing was seen as less desirable than a pattern of testing that would identify clearly a representative sample of students to whom a stipend would be paid for their cooperation. In contrast, the voluntary participation without stipend by third-quarter seniors in the Honors College seemed to reflect the most optimal conditions for assessing performance on the several tests. An institution that proposes to measure generic skills at the close of the undergraduate experience would be advised to stipulate such participation by students in graduation requirements; otherwise, the sampling aspect will continue to present problems.

One of the unexpected outcomes of the APEP experience at Ball State University was the increased sensitivity to the need for determining in what areas students might acquire the skills under study. The faculty expressed a particular concern for identifying the intentional steps instructors take to reinforce skill development. This expression was seen most obviously in the work of the General Studies Committee. During the period of APEP evaluation the committee, many of whose members were active in the project, began an intensive review of the program of general education. Its dialogue with departmental representatives, examination of course syllabi, and other program review activities led the committee and APEP participants to conclude that insufficient attention was being given course objectives and instructional practices bearing on the development of intellectual skills. Apart from the information provided on student and program performance in relation to the four generic skills under study, this recognition of the need for curriculum improvement may have been the single most important outcome of the study.

Recommendations

On the basis of the experience given a relatively large sample of the faculty and staff in the past two years, it seems appropriate to recommend the following:

- that the Undergraduate Educational Policies Council develop a policy statement that commits the university to a continuing program of assessment of student performance in the skills of communication, quantification, analysis, and synthesis, with special attention to entrance and exit level of skill performance
- that the university officers determine which administrative unit should serve as an assessment center and provide the locus of responsibility and authority for a continuing program of student assessment, and that they provide for such a center the cooperative use of necessary resources such as specialists in research design, psychometrics, computerization, and data analysis
- that the General Studies Committee in its current planning for a revised program of general education give special attention to the need for greater structure in the program; course approval criteria that reflect the importance of course objectives and instructional strategies related to generic skills; and continuous evaluation of the program
- that both the Department of Educational Psychology and the Department of Secondary, Higher, and Foundations of Education encourage as a component of their graduate studies increased research of problems, issues, and practices related to undergraduate instruction and curriculum development at this institution
- that the Office of Instructional and Professional Development provide, through workshops and other means, specific assistance in developing course objectives and instructional strategies related to generic skills.

Written By

Thomas A. Kaluzynski and Victor B. Lawhead

Institutional Contacts

Thomas A. Kaluzynski
Victor B. Lawhead

California State University, Chico

Institutional Profile

California State University, Chico began in 1887 as Northern Branch State Normal School of California and is the second-oldest school in the current California State University system. Its purpose was "the training and education of teachers in the art of instructing and governing public schools of the state." As might be expected, Chico State has changed dramatically from its founding status of five faculty members and ninety students. Other colleges and universities began to accept appropriate work in liberal arts courses at Chico for transfer credit in 1913. Chico State Teachers College was first authorized to confer the Bachelor of Arts degree in 1924. Master's degrees were first offered in 1949, and in 1971 the institution was formally titled "university."

At the present time, Chico State can most accurately be described as a large liberal arts university whose primary mission is instruction. A wide variety of other activities, ranging from basic research through community service also characterize the university's operations today.

Although it is risky to predict the future, it seems fairly certain that the university's future mission will be less broad than that pursued in its past. Without major responsibility for either two-year programs or research functions, it is anticipated that Chico State will grow in a variety of directions addressing the needs of California's population and developing in a manner commensurate with new understandings of educational structure and process. No future development is expected to change either the basic instructional mission of the university or the faculty commitment to instructional excellence.

Brief statements cannot fully describe California State University, Chico, but the following characterizations will certainly provide the reader with a good overview:

- a geographic service area of twelve northeastern California counties, which is about one-fifth of the total area of California, and larger than ten of the fifty states
- a remarkably beautiful campus of more than 100 acres, augmented by an 800-acre university farm and numerous agreements providing for the conduct of university instruction at a variety of off-campus sites throughout its service area
- a faculty of about 700, the majority holding doctoral degrees, with a similar-sized support staff; of the latter group, about 45 percent working in direct support of instruction (instructional, administration, library, and so forth)
- a curriculum covering most of the major academic subdivisions of knowledge. Only architecture, law, library science, and theology are absent. Baccalaureate degrees are awarded in more than fifty areas of study, and about half of these offer master's degree programs as well
- a student body of about 14,000. Of these, more than 80 percent attend full time. More than 85 percent are undergraduates. The sexes are equally represented. About 40 percent of the university's enrollment is new to Chico State each year, and of these, about half are transfer students from other (mostly two-year) colleges and universities.
- few students are academically disqualified--perhaps two or three percent each year--and relatively few students migrate to other colleges and universities. About 70 percent of our freshmen return to continue their education in the sophomore year.

In 1973, the former Federation of Regional Accrediting Commissions of Higher Education received a grant from the Danforth Foundation for a three-year study, "Looking Toward the Development of New Techniques for the Evaluation of Institutions of Postsecondary Education Which Would Attempt to Emphasize Outcomes Measurements." Chico State was a participant in that study, and a description of its involvement and accomplishments is reported (under the pseudonym of Big Creek University) in *Evaluation of Institutions of Postsecondary Education: Assessment in Terms of Outcome Through Institutional Self-Study.* Chico State gained much valuable experience from that study, of use in improving both its educational planning and ability to conduct self-appraisals.

Background

In addition to the institution's involvement in the self-study experimentation described earlier, Chico State has a unique history of evolution in its general studies program. Beginning in 1972, the faculty, in partnership with the academic administration, addressed the issue of establishing general studies goals and building a program based on those goals. After establishing criteria that supplemented the goals, an ad hoc committee created by the faculty senate spent three years reviewing every course then listed as satisfying general education requirements. They met with faculty members and school administrators and formulated a new list of recommended courses. In the end, the general education list was pared from some 2,000 to 250 courses. Although there was understandably some unhappiness with individual decisions, there was also a general acceptance by the faculty and administration that the job had been well done and that the university had the beginnings of a sound general education program. The ad hoc committee was replaced by a faculty advisory committee, and the undergraduate dean in the office of the vice president for academic affairs assumed the role of general education coordinator.

The general education program has consistently been addressed in the university's planning process, and it is generally recognized that the program must be protected from course proliferation and narrow departmental resource concerns. The faculty senate has been highly protective of the program and consistently resists efforts to turn general education into a political issue.

Chico State became involved in the APEP program at a point in its history that was both fortuitous and problematic. The university had, up to that point, seven years of experience in developing a general education program in terms of a set of goals generated and approved by the faculty. Many faculty members however, felt that some revisions were in order and a new phase of program building was appropriate. Furthermore, a modest process evaluation of the existing program had led to additional questions about the impact of general education, as measured by differences in the levels of accomplishment on achievement tests of native and transfer students as well as of freshmen and seniors.

During this period a California State University system task force was working to develop a new proposal for the California Administrative Code segments on general education. University officials were aware that the trustees would act on this matter by May 1980 and subsequent to that action the campuses would have approximately one year to make their programs fit the new guidelines. Drafts of the system's task force materials were circulated for campus response, and it became clear that the new guidelines would be far more prescriptive and detailed than ever before. The proposal for participation in the APEP project was written in this context and placed major emphasis on gathering data from students who would complete the existing general education program. In this way, data would presumably be available for drawing comparisons between their accomplishments and those of students completing the program to be initiated in the fall of 1981. In addition, freshmen/senior differences and the differences between native and transfer students would be measured.

Implementation

Research Design

The research design for the study is given below in the form of four basic questions asked about any combination of goals for which measurements were obtained. The narrative makes references to transfer and native students. "Native students" are defined as those freshmen entering Chico State with no transfer work from another college or university. "Transfer students" are those with fifty-six or more units from other colleges or universities accepted at Chico State. Students with intermediate amounts of transfer work were not incorporated in the design.

What is the level of accomplishment of Chico State students?

Two groups of thirty randomly selected native seniors were tested in the fall semesters of 1980 and 1983. Some scores of these groups compared with national, state, or local norms for some of the tests being used.

Is experience with general studies and other courses at Chico State correlated with changes in the level of accomplishment?

An additional group of 100 randomly selected students who had been tested as freshmen in the fall of 1980 were tested again in the fall of 1983.

The above question was answered by comparing the scores of the 100 seniors with their same scores as freshmen. Freshmen-versus-senior differences were assessed both within 1980 and within 1983 to detect how students change during their college years.

When do changes in the level of accomplishment take place during a student's career at Chico State?

A separate, randomly selected group of thirty native freshmen were tested each fall semester from 1980 through 1983.

A comparison of these four groups of scores was intended to provide information relevant to detecting the point in time at which changes occur and the effects of repeated testing. Unfortunately, student participation problems reduced the sample sizes, and attention to this question was deleted from the study.

Is there a difference in the level of accomplishment of native and transfer students?

Two additional groups of thirty randomly selected transfer seniors were tested--one in the fall of 1980 and the other in the fall of 1983. The scores of these transfer senior students were compared with the scores of native seniors.

The Role of APEP in General Education Program Development

Faculty on our campus were aware that by May 1980 there would be new system guidelines for general education, and that there would be about one year to develop local implementation policy. Clearly, neither the Educational Policies Committee of the Academic Senate nor the General Studies Advisory Committee was willing to spend any time on any revision or rethinking of the general studies program until the system trustees had acted. We of the project staff knew that once this radical redistribution of general education content was announced, campus interest and participation in formulating local policy would be intense and pervasive. At this time, we made several proposals to the president and to the vice president for academic affairs. All the proposals, listed below, were subsequently accepted.

- In the fall of 1980 we would construct an interim program for the 1981-82 academic year. This interim program would provide time to develop a more thorough approach to revising our policy;
- The 1980-81 academic year would be devoted to a thorough and careful restudy of campus general education goals in the light of system policy and the development of local policy for program building, monitoring, and evaluation, as well as criteria for course review;
- The APEP generic skills would be systematically embedded into the new criteria;
- All courses proposed for the new program would be carefully and individually reviewed by a faculty committee using course descriptions, sample examinations, and sample assignments, as well as criteria for course review;
- By the end of the fall semester 1981 all but the upper-division program would be in place and by April 1982 the whole program would be complete;
- Every five years one-fifth of the programs would be rebuilt *de novo*, with each course reviewed in terms of the general studies criteria. We realized that this timetable was out of harmony with the sequence of the APEP paradigm as it evolved.

At that point, we had to work with the existing general studies goal statement as our conceptual definition. It was decided to use a two-committee structure. The governing committee for the APEP project was the long-standing advisory committee on general studies, whose charge was to advise the vice president for academic affairs, or his designee, the undergraduate dean. This committee was to have the technical and advisory support of a committee whose members were faculty members with expertise in tests and evaluation and who came from a variety of academic areas. The university's test officer and the director of institutional research served as staff to that committee.

- Student accomplishments in general studies be regularly measured and the outcome information provided to the appropriate committees. Using the criteria stated in our general studies goals and those of APEP, the evaluation committee prepared a series of proposals for review by the General Studies Advisory Committee. After several consultations and subsequent revisions, a final list of evaluative instruments was selected.

Test Measures

All student participants took a thirty-minute survey titled the *Student Information Form* (SIF; Cooperative Institutional Research Program). This survey provided demographic and attitudinal information valuable for later analyses of the outcome measures. All students also had a university record of their grade-point average, workload, etc., for each semester under study. The remaining measurements are described in sections identified by their focus on local or APEP goals or both:

I. Literacy (CSU, Chico's Goals of General Studies); Communication (APEP); and Quantification (APEP)

A. *A Student Composition* writing task constructed locally. This was a sample written under "compare and contrast" instructions such as those used in the state's English Placement Test and English Equivalency Examination. Local holistic scoring was conducted by experienced readers from the English Department faculty, and the results were compared with state norms for students in the CSU system.

B. The *Cooperative English Test* published by ETS. Two sub-scores titled Reading Comprehension and English Expression were obtained from this standardized, objective test having national norms.

C. *Measures for Academic Competencies in General Education* available from Jonathan Warren of ETS. There were two of these tests, one in communication and the other in mathematics.

D. *National Assessment of Educational Progress (NAEP): Mathematics* published by the U.S. Office of Education. This test required students to use numbers, presented in digital, figurative, and narrative form, in problem-solving tasks. National norms were available for 17-year olds and for young adults (18 to 35) but not for college students.

II. Critical Reasoning (CSU, Chico Goals of General Studies); Problem-solving (CSU, Chico Goals of GS); Analysis (APEP); and Synthesis (APEP)

A. *Test of Thematic Analysis* published by McBer. It asked students to identify and describe the differences between two groups of stories. High scores represented a differentiation and evaluation of story content; low scores represented differentiation based more on the reader's personal response.

B. *Analysis of Argument Test* published by McBer. Students were asked to produce arguments on both sides of an issue. Scores represented the ability to organize and evaluate arguments about both sides of an issue.

C. Rokeach *Dogmatism Scale*. This test was not published but exists in the research literature of social science. The test was designed to assess the openness of an individual's belief system.

D. *Measures for Academic Competencies in General Education* published by Jonathan Warren of ETS. There were two of these tests, one in analysis and the other in synthesis.

III. Artistic Expression (CSU, Chico Goals of GS)

The College Level Examination Program (CLEP) General Examination iin Humanities) published by ETS. Subtest scores in Fine Arts and Literature were obtained and scored in reference to national norms.

IV. Form of the world in which we live (CSU, Chico Goals of GS)

The *CLEP General Examination in Natural Science* published by ETS. Subtest scores in physical science and biological science were obtained and scored in reference to national norms.

V. History and Government (CSU, Chico Goals of GS)

The *CLEP General Examination in Social Science-History* published by ETS. Subtest scores in social science and history were obtained and scored in reference to national norms.

VI. Aspirations and problems of subcultures (CSU, Chico Goals of GS)

No tests were proposed for this goal.

VII. Influence of technology (CSU, Chico Goals of GS)

No tests were proposed for this goal.

VIII. Integration of major work (CSU, Chico Goals of GS)

No tests were proposed for this goal.

IX. Valuation (APEP)

AVL Study of Values published by Houghtin Miflin. Six basic interests were assessed with this instrument: theoretical, economic, aesthetic, social, political, and religious. National norms were available.

The above list was the result of substantial effort. Faculty committees reviewed a great amount of pertinent literature, met regularly, and debated and discussed the issues with care. They found it to be an extraordinarily complex task and one wherein no single result could be defined without some ambiguity. Despite structural difficulties, the proposed design was submitted with the confidence that it represented substantial progress toward several ends, namely:

- This proposal defines several ways in which better information on Chico State's accomplishment in general education can be acquired;
- It was consonant with APEP's orignial definition of project tasks;
- It was consonant with the history and current conduct of self-study related to general education at Chico State and at the CSU system;
- The information acquired from this study would stimulate greater and continuing faculty involvement and dialogue in matters related to general education.

The testing was completed on schedule. The participation rates of students were lower than we had hoped even though sample students were individually contacted and paid for their time. However, the first phase of the testing program was completed with adequate numbers and the baseline data needed were generated. That done, we returned to stage one as, in the late fall 1980, the faculty was now ready to address the general studies criteria in response to the system-mandated guidelines.

Campus Response to New System Guidelines for General Studies

As mentioned above, the Educational Policies Committee of the faculty senate was charged with developing a policy recommendation for the entire academic senate to consider. This strategy included receiving proposals from the various academic units on the campus with ideas for criteria. A subcommittee of the Educational Policies Committee developed early drafts of some of the broad conceptual statements. The Educational Policies Committee as a whole reviewed these conceptual statements as well as the more specific criteria. The project was given the broadest possible publicity, with people from the various academic areas invited to attend the meetings as proposals pertinent to them were discussed.

Our intention was that the process be open and the widest possible participation be solicited because we knew that the redistribution of general studies was going to have a significant impact on all units of the campus. Open hearings were held and subcommittees wrote policy to incorporate conceptual statements and explicit criteria in their work. Many revisions, discussions, compromises, and new documents emerged. A draft policy document was sent the faculty senate and was also distributed widely in the campus community. Sessions of the faculty senate were well attended while the new general education policy was debated, and in May 1981, with relatively few revisions, it was accepted by the president.

The skills for APEP, as well as other local goals, were written into the policy. Fall of 1981 was devoted to the selection of courses for the program by faculty review committees established under the policy. The total number of courses was limited to 200, so the issue was not merely which courses met the criteria but, rather, which courses best met them. The burden was on faculty members to demonstrate that not only did their courses cover the content area, but also communication, particularly writing skills (in all courses), quantification (where appropriate), critical thinking, analysis, and synthesis. In the upper-division program, course proposals had also to demonstrate an integrative component and a focus on values. Examinations and assigments for all courses had to be consistent with course outlines.

The lower-division program is now complete and the upper-division program was completed in April 1982. Our next task is to review the testing program to determine whether its content is still consonant with the program criteria. The evaluation committee will be reconvened for that task. The next large testing by Institutional Research with standardized instruments will have taken place by 1983-84. The previously tested freshmen, now seniors, will have completed the old general studies program; other freshmen will have begun under the new one. Ultimately, a useful evaluation of student achievement changes will result. These changes can be assessed for both pre- and post-APEP times and will likely be the stimulus for a major review or progress report on general studies in the mid-1980s.

The Present Impact of APEP

The impact of APEP at Chico must be viewed in relation to other attributes of our history and plans. APEP merely augmented what was on going and planned. APEP, at Chico, was definitely not a separate endeavor with well-defined initiation and termination dates. Chico has been paying special attention to APEP-like topics for more than a decade and intends to continue that attention. APEP, per se, provided welcome help, additional rationale, and other benefits to this work. A brief describption of that directly related to APEP follows.

One result to date is the acquisition of information related to the level of measured achievements by Chico State students. In general, Chico's results were simultaneously informative, reassuring, and insufficient. We have been able to compare some of our test results to those published as norms in test manuals or to study results such as those reported in *The Impact of College on Students* by Feldman and Newcomb. We have not felt confident, however, that we had access to information representative of college student achievements in AASCU institutions or any other well-defined population. The problem is mostly that such information does not exist.

Representative information could not exist at Chico, where almost half of a well-defined sample of students declined to participate. Also, information representative of national college student populations is usually absent and often characterized by sampling flaws when present.

A second result to date is the acquisition of information related to the mean achievements of Chico student subgroups. We expected to find differences between our fall 1980 freshmen and seniors. We were curious about the differences between native and transfer seniors in fall 1980 and in spring 1984 replicate. In general, the results to date have been informative and somewhat reassuring. No other APEP institution reported finding significant group differences in the hypothesized direction on all the tests used, and Chico State also could not report such a result. Test information such as that described here is less vulnerable to the sampling criticisms previously mentioned. The absence of defensible norms is almost irrelevant, and the bias resulting from incomplete local participation is somewhat amenable to additional analysis.

A third result to date involved some of the APEP tests on written communication and mathematics. The communication test results were compared with those of a similar study in 1975 and helped convince the faculty committee on writing that more stringent measures were needed. The committee recommended, and the president agreed, that all junior local students, native and transfer, had to take a screening test in written English. Those not passing at an acceptable level would not be permitted to continue their upper-division writing work until the deficits had been remedied. This policy began in the fall of 1982. Similar attention is being devoted to mathematics. A fourth result to date was the acquisition of program portrayal information of two kinds. A survey of general-education class students and their instructors produced a description of the amount of attention given to each general education goal in classes satisfying general education requirements. We found some classes neglecting some goals and some discrepancies between instructor descriptions of what was delivered and student descriptions of what was received. This informative feedback generated some valuable dialogue.

The second set of information produced has yet to be published. This is a description of the amount of different kinds of course work taken by students in different majors. Those familiar with the literature of higher education management systems may recognize its name: "Induced Course Load Matrix" (ICLM). This information describes the impact of a university's new general education policies on the curriculum consumption of its student body. It also reports student surrogate outcomes from the perspective of quantity of schooling. These course consumption or ICLM reports will be regularly published starting with the fall semester 1983. Historical reports will also be available at the same time.

Some of the overall results described above depended on completion of the testing facet initiated in fall 1980. Other aspects of Chico's program use that information but are not dependent on it. One such program is the University Planning Data Base (UPDB). There we have been working with a computer data-base management system called Scientific Information Retrieval (SIR Inc., Skokie, Illinois), which has great promise for managing the work of a university intending to develop an analytical management information system. The UPDB should be fairly functional by the end of the year. It will describe resource utilization and outcomes, by program, in a manner that is highly informative and exceptionally easy to use. Chico's involvement in the APEP project has helped bring student outcome information into the UPDB, and this we regard as particularly meritorious.

Conclusions

APEP has proceeded, albeit at a snail's pace, between cessation of active supervision by AASCU and the present time. The UPDB is regularly publishing educational statistics for the university community, and post-tests have been administered to those original APEP students willing to participate. Student outcome measurements are now obtained annually. Approximately one-fifth of each year's seniors are sampled and offered both R. Paces survey measuring the "Quality of Student Effort (Higher Education Research Institute, UCLA) and a locally developed survey of "Educational Satisfaction." All degree recipients are offered a "Survey of Recent Graduates" about four months after their graduation. Many post-APEP research reports are still possible and still intended. They will be published as time for them is available to the institutional research office.

The advent of APEP and Chico's involvement in the project continues to promise a unique benefit to the university. The work and faculty involvement associated with APEP succeeded in "defining" a stand-alone operation that provides informative feedback on the university's most important goal, student achievement(s). Its cost was modest--about 1/1000 of the institution's annual budget--and its nature sufficiently flexible to accommodate all except those opposed to any form of standardized test-based appraisal. We propose conducting the APEP operation every fourth year in the foreseeable future. The result, if implemented, will add a history of student accomplishments to our history of institutional program and policy changes.

In retrospect, there were several important decision points in this project once it was underway. The first was to go ahead with a testing program based on the definitions we had rather than develop new ones in the context within which we were working. The second was to capitalize on the campuswide interest in general education as a result of system mandates to embed APEP skills into the new program and concomitantly to delineate the relationship between skills and content. Finally, we took the opportunity to build an evaluation component into the new campus policy and related it to subsequent program building, thus ensuring that the spirit of APEP goals and processes have been institutionalized.

Primarily the results will be used for evaluating and pointing out needed revisions to our general education program. This information will be provided to deans of schools seeking participation in the program in general education, so that they may be aware of strengths and weaknesses the testing program has identified. For the public we can provide assurance that a systematic program evaluation project is in place, that it has an impact, and that program changes will be made where weaknessess are identified. We will be satisfied with the program only when all of the freshmen/ senior differences are apparent and significant.

We do not plan to release the results of the testing either to celebrate accomplishments or to deplore weaknesses. This information is to be used internally for educational purposes; to monitor, to point out areas requiring change, and to evaluate the effectiveness of our decisions. "Outsiders" with a need to know specific results will be allowed access to such information. (Contact the director of the Office of Institutional Research.)

Written By

Patricia Brose and John G. Safarik

Institutional Contacts

Robin Wilson, president
Patricia Brose, professor of education
John G. Safarik, director of institutional research

North Adams State College (CO)

Institutional Profile

Located in North Adams, a city of 17,000 in the northwestern corner of Massachusetts, North Adams State College (NASC) is one of ten state colleges in the Massachusetts System of Public Higher Education. Chartered in 1894, it currently enrolls approximately 2,100 FTE daytime students mostly from Massachusetts but also representing fifteen states and five foreign countries. North Adams State College is primarily a residential college with approximately 80 percent of its students living on or near the campus.

The college offers fifteen majors and eleven minors, as well as a variety of special programs. Approximately 30 percent of the students are enrolled as business majors and 10 percent each in teacher preparation programs, sociology, and computer science. Other major programs are in biology, chemistry, English, history, interdisciplinary studies, mathematics, medical technology, philosophy, physics, and psychology.

Background

In the fall of 1977, North Adams State College instituted a new general education curriculum to replace a purely distributive one. The new curriculum established seven general education categories, defined the purpose of each category, and identified specific courses for each category. During the next two years, the curriculum committee monitored the program, made decisions about what new courses should be allowed into the curriculum, and undertook a limited evaluation of the program in the spring of 1979. There was general agreement that a more rigorous evaluation of the curriculum would be desirable.

It was in this context that the college first heard of the Academic Program Evaluation Project. James R. Roach, vice president for academic affairs, submitted a miniproposal to AASCU, and the college was notified in August 1979 that it would be one of ten colleges and universities participating in the project.

Implementation

Preliminary Phase

Several issues were paramount during this early period of the project. What would be the best way to evaluate the general education curriculum? How could an evaluation be carried out so that the faculty would be most supportive? To what extent should the APEP guidelines drive the program evaluation? In general, there was agreement about the need to define each category of the general education curriculum more carefully, that broad faculty participation was very important in this process, and that the college should use the APEP guidelines as a resource rather than be driven by them.

At a faculty meeting in September 1979, it was announced that the college would be participating in the APEP. Few details were provided at that time, but generally the message was that the project would be of assistance in evaluating the new general education curriculum.

Following receipt of some start-up guidelines from the national office in mid-October, the campus director and coordinator of the project discussed alternative strategies for getting the project underway at NASC. the college's incentive awards committee, in cooperation with the curriculum committee, was planning a "Day of Conversation" to involve the faculty and administration in discussing and clarifying the meaning of general education. It was decided to capitalize on this event and to consider it the beginning activity of the project.

Approximately thirty-five of the total faculty of 100 attended this Day of Conversation. After a few opening remarks, faculty members formed interdisciplinary small groups to discuss such questions as: What is your definition of general education? What do you think it means to be an educated person? What intellectual abilities should a student be able to perform as a result of a good general education? What criteria should the curriculum committee be using when deciding whether a particular course should be accepted as a general education course?

Once groups had finished their discussion, group leaders took part in a panel presentation of the findings and fielded questions from the floor. The final segment of the conference was an address by the president, who spoke in favor of general education and identified a number of problems and possibilities related to general education. Reference was made to APEP and the role it could play in helping the college improve the general education curriculum.

The curriculum committee was given overall responsibility for overseeing the project on campus. The primary tasks for the first year were identified as follows:

- Elaborate on and clarify the objectives of each category of the general education curriculum.
- Define the objectives in terms that will allow faculty members to determine when they are being addressed.
- Conduct an assessment of existing general education courses to determine when the objectives are being addressed.
- Take steps designed to increase the extent to which the objectives are being addressed.
- Begin work on an evaluation design to be implemented during the second year of the project.

In discussing their charge, committee members agreed that it probably would be wise to downplay any mention of a general evaluation effort during the early part of the project and to speak in terms of the project's helping the faculty do a better job of helping students achieve the objectives of general education. It was decided not to take any further steps with the project before the national meeting of all project participants, scheduled for December 1979.

Conceptual Definitions

The national meeting marked the first turning point for the project at NASC. Prior to attending the conference, the general plan had been to make more specific the objectives of each general education category as a foundation for subsequent assessment and evaluation and for APEP to constitute one part of this larger effort. After the Washington meeting, the broader effort gradually was postponed and attention focused on defining more clearly the general education abilities of communication, quantification, analysis, and synthesis.

The decision to concentrate on these intellectual abilities rather than attempting to define specifically all the concepts and abilities associated with the various general education categories was made for several reasons. Simple time constraints were a primary consideration. Also, there was a feeling that the college should fulfill its obligations to the national project. A strategy was needed that would allow focusing on the four intellectual abilities identified by APEP without losing sight of broader local objectives. Also, the selected strategy would have to ensure work on conceptual definitions did not become an arid intellectual exercise and involve a significant number of the faculty.

Four members of the faculty were commissioned to write working papers on the four intellectual abilities of communication, quantification, analysis, and synthesis. These working papers were to examine literature on the abilities and to describe the various ways the abilities are defined. All members of the faculty and administration received a memorandum summarizing the background of the project, describing progress to date, and inviting them to attend a meeting to discuss the four papers.

Over a quarter of the faculty attended this meeting. The authors of the papers made brief presentations, interdisciplinary subgroups discussed the papers, and group leaders subsequently reported on reactions to the papers. Revised papers were distributed to the entire faculty for review and comment. The revised papers described the conceptual territory for each of the four intellectual abilities but did not attempt concise definitions. Was this sufficient as a foundation for faculty members to work on operational definitions within their own disciplines? Would anything be gained by abstracting from the richness of the papers a concise conceptual definition and then having to move back to more concrete behavioral statements?

The curriculum committee decided against formulating tight conceptual definitions. Instead, outlines of the four papers were taken to Washington and guidance sought on how best to proceed. at the Washington meeting it became clear that some institutions had prepared tight conceptual definitions, others had listed subcomponents of the intellectual abilities, and other institutions had done both. It became clear that North Adams would need to move quickly to formulate operational definitions from its conceptual descriptions.

Feeling the pressure of time, not wishing to lose momentum on the project, and desiring to involve the faculty in operations as soon as possible, the project committee decided to draw from the list of subcomponents that had been developed by other participating institutions those that were most clearly stated. Thus, North Adams State moved from its four papers describing the rich conceptual territory associated with each of the intellectual abilities to lists of subcomponents for each of those abilities. These were to be the reference points for operation.

Operational Definitions

The need to move from conceptual discussions to operational definitions was clearly recognized. Faculty members would not be likely to sustain their involvement unless they saw some direct implications of the project for their own work in the classroom. At the same time, it was important to keep in mind the purpose of moving toward an evaluation of the general education curriculum. What strategy would be most effective?

One possibility was to bring together faculty members teaching courses in the various general education categories, group them by categories, and ask them to work on operational definitions for communication, analysis, synthesis, and quantification. Given such definitions, an instrument could be devised or selected for evaluating student performance across the entire program.

The turning point for the entire project at North Adams State came when it was decided that the best approach to an eventual evaluation of the general education curriculum was to begin by having individual faculty members work within the context of their own courses to improve student performance of the intellectual abilities. The rationale was that if faculty members were to get involved in a systematic assessment of student performance within their own courses they would be more receptive to a systematic evaluation of the entire general education program. The project would later include some beginning efforts at programwide evaluation but primary emphasis would be given to work at the level of courses.

With this strategy in mind, a memorandum was sent to all faculty members inviting them to make a commitment to the remainder of the project and spelling out what would be involved. Each participating faculty member would focus on a course he or she taught and, in conference settings, work with three to five others in support groups to:

- share the details of their courses as currently taught
- identify specific ways students might communicate, quantify, analyze, and synthesize
- explore ways to foster more effective intellectual functioning in the content of each course
- design methods for evaluating student performance of the intellectual abilities.

Over twenty faculty members from most of the disciplines volunteered to take part.

Development of Assessment Tools

The APEP paradigm developed by consultants to the project provided a common frame of reference for participating institutions. Although North Adams State had chosen to work first within the context of individual courses, it was important that work proceed with reference to the paradigm. This meant that participating faculty members would need to develop assessment items and criteria for their own classes, assess student performance and factors affecting that performance, make judgments about student performance and their course, and decide on possible changes.

Accordingly, two meetings were held with volunteer faculty members in May 1980 and a two-day retreat followed in June. The first meeting was viewed as an orientation session. Each participant was provided with a notebook containing key documents from the central office and memoranda and papers that had been developed locally. Expectations and ground rules for the project were made clear and questions answered. Throughout this first meeting, the emphasis was on collegiality.

The first working session was held in late May 1980, and the objective was simply to get faculty members talking about their own courses and ways that they have sought to challenge students to communicate, analyze, synthesize, and quantify. Each participant brought copies of materials for the course he or she had selected and, in interdisciplinary small groups, shared with others the objectives for the course, instructional strategies commonly used, and evaluation modes. Later in the session, each group focused on one of the intellectual abilities and discussed specific ways they sought to foster that ability within their several courses. By the end of this first meeting, faculty members had become comfortable talking with one another about their courses and about ways of fostering higher-level thinking. They especially felt good about talking with faculty members from other disciplines and expressed a willingness to participate in the two-day retreat planned for mid-June.

This retreat was seen as critical to the future success of the project. It would represent the first time faculty members would be asked to look analytically at their own practices and to consider ways of writing more effective assessment items. Prior to the retreat, all participants were sent a "Homework Packet" asking them to prepare an outline of their course and, on a separate sheet containing all the subskills of the several intellectual abilities, to indicate whether each subskill was one they would like to emphasize more explicitly or whether it seemed irrelevant to the course. The intention was for each faculty member to arrive at the conference having thought about the various intellectual abilities in detail.

The basic strategy for this workshop was to engage the participants in the writing of assessment items geared to a subcomponent of a selected intellectual ability and to discuss criteria for judging student performance on the item. Each subgroup identified a single component of an intellectual ability, determined some possible activities that would call on students to perform that subskill, and then designed an actual assignment (complete with explicit instructions) to be shared later with all participants.

The next morning, each subgroup continued to focus on the assignment they had designed but this time attention was directed to anticipating the range of performances and on developing criteria they would use to discriminate among these performances. In the process, several groups found it necessary to rework the wording of their assignment in order to make it clearer.

These exercises in operationalizing were interspersed with discussion sessions. For example, many found it difficult to concentrate on a single component of an intellectual ability but felt less uneasy when it was acknowledged that probably many intellectual operations would be involved in any given assignment and that they were being asked simply to give most attention to a particular component. Large-group discussions of such issues and the informal socializing that occurred during the cocktail hour, at meal times, and the extended nightcap session all contributed to the growing identity of the group. Again and again, participants acknowledged the value of having a chance to talk about teaching and learning with those representing other disciplines.

Everything that had been done during the first year of the project can be seen as being preliminary to really going operational. Conceptual descriptions had been written, subcomponents identified, and the faculty acquainted with what is involved in preparing assessment items and criteria. During the next year of the project, each faculty members would build on this foundation, applying what they had learned to one of their own courses.

To sustain momentum, participants were sent a summary of the conference, a draft plan of activities for the next year, and a resource packet to aid in their individual planning. Each participant was encouraged to establish clearly the objectives for the course they had selected, concentrate on one or a few components of an intellectual ability, and be imaginative in designing the assessment items. An overview of different assessment modes and their advantages and disadvantages was provided. Finally, suggestions were given that would assist the faculty in specifying criteria for judging student performance levels.

Throughout the 1980-1981 academic year, participants tried out their assessment items, refined them, and tried them again. This process was facilitated by occasional meetings with other members of their support groups, by a progress-sharing conference in October, and by a mid-year conference. Most faculty members chose to concentrate primarily on one or another subcomponent of analysis, whereas others concentrated on communication, synthesis, or valuing. Some faculty members were very systematic in their experimentation; others found it difficult to keep their focus on a particular intellectual subcomponent. Again, a real benefit of involvement was the interaction among representatives of different disciplines.

The APEP paradigm being developed and refined during that year called not only for the preparation of assessment items and criteria for judging performance but also for faculty members to consider factors that might affect student performance. Informally, participants had discussed such factors early in the project, when they shared with one another ways they sought to foster higher-level thinking. This stage asked them to be more rigorous in identifying such factors and in relating them to the performance on assessment items they had designed.

During the two-day conference in January 1981, small groups consisting of an interdisciplinary mix of faculty members were asked to list teacher and student behaviors they felt helped students analyze, communicate, quantify, and synthesize. Each group concentrated on a particular intellectual ability and the list of items was then shared with all participants.

With this discussion as background, each faculty member was asked to think about his or her course in detail and to indicate the means they were using to evaluate student performance (essays, journal, class participation, etc.), to list any of their behaviors that might foster intellectual ability, and to identify student behaviors that might affect their performance.

By the end of the 1980-1981 academic year, each faculty member had designed and refined assessment items for his or her course, considered criteria for judging student performance on those items, and identified teacher and student behaviors that might affect performance.

Program Assessment

While much of the 1980-1981 academic year was devoted to faculty members working within the context of their own courses, steps also were being taken to move toward an evaluation of the entire general education program. At a meeting of participants in October 1980, sample assessment items from the Educational Testing Services' measures for general education and from McBer and Company's *Analysis of Argument Test* were duplicated for examination. Faculty members answered each of the items and were then asked to consider how well the items measured analysis, communication, quantification, or synthesis as defined by the subcomponents identified the previous spring.

Reactions were largely negative. Most of the discussion focused on the ETS items. There were differences of opinion on the "correct" answers. The wording of the questions was criticized. Many expressed doubt that multiple-choice items were appropriate for measuring high-level intellectual abilities.

In spite of these reactions, participants were asked to administer at least one of the ETS tests to their classes prior to the end of the semester. It was suggested that they might find it interesting to correlate students' results on whichever test they selected with students' performance on local assessment items they had designed. Approximately half of the participants agreed to administer at least one of the ETS tests. The results were scored and returned but for a variety of reasons faculty members did not explore correlations.

Paralleling the experimentation with ETS items was exploration of the instruments developed by the ACT's College Outcome Measures Project. All members of the steering committee attended a two-day workshop in Albany, New York, and came away impressed with what they had experienced. At this point of the project, however, the primary emphasis was still on faculty members developing course-specific assessment items, so further consideration of ACT/COMP tests was postponed.

A further step to gather programwide information was undertaken during the January 1981 two-day conference described earlier. Participants selected four items each for analysis, synthesis, communication, and quantification from the most reliable forms of the ETS measures for general education. The resulting sixteen items were assembled into a composite test for measuring high-level thinking abilities. This test was administered to 482 students in February 1981. Later in the spring, 338 of these same students took another form of the examination, thus providing the college with at least some information on student performance in all majors and classes.

To gather some program-portrayal data, it was decided to try to identify the SAT scores for each student taking the composite examinations and to ask faculty members to include the following two items on a questionnaire for student evaluations of courses and instructors:

- "This course required me to think, not just memorize"
- "The instructor provided constructive written or oral comments on students' work."

The results of these programwide efforts, together with the data gathered on course-specific items and program portrayal factors, constituted North Adams State College's work on the latter stages of the APEP paradigm.

Outcomes

Because of North Adams State's two-pronged approach, data were generated at both the course and program levels. From the beginning of the project, faculty participants had been assured they would maintain control of any data generated by their course-specific assessment items and that they would not be required to share results with the community at large. However, they were invited to write individual case histories of their experience with the project and to reflect on changes that might be in order. A majority of the participants did prepare case histories, some more detailed than others. All of them, however, commented favorably about their project experience. Most described the kinds of assessment items they had designed and the procedures they had used for stimulating higher-level thinking among students. Some of the respondents commented on how the project might be continued; all of them felt one of the primary benefits had been the involvement with faculty members from other disciplines.

At the program level, data were generated primarily by the two composite tests administered in spring 1981. In general, the results of both tests revealed clearly that seniors performed better than juniors, juniors better than sophomores, and sophomores better than freshmen. The following tables show the percentage of correct responses for each segment of the composite tests:

First Test	*Number*	*Analysis*	*Synthesis*	*Mathematics*	*Communication*	*Overall*
Class '81	78	58	78	71	69	67
Class '82	110	53	76	65	54	61
Class '83	104	45	70	60	55	56
Class '84	190	43	61	57	41	49

Second Test	*Number*	*Analysis*	*Synthesis*	*Mathematics*	*Communication*	*Overall*
Class '81	60	63	51	63	61	57
Class '82	75	67	43	66	53	55
Class '83	75	54	44	61	47	49
Class '84	128	52	40	59	38	44

It was commonly agreed that the items on the second composite test were either more difficult or less clear than those on the first test. Also, many of the students expressed their impression that they "had already taken" this test and therefore had not taken it seriously.

Unfortunately, neither of the attempts to gather programwide information on factors that might affect student performance were successful. SAT information on students taking the composite tests was not readily available, and it was decided not to carry out a record-by-record search. Second, only a few faculty members chose to include on the Student Instructional Rating Form the two questions designed to elicit student perceptions of the challenge of the course and the amount of instructor feedback.

To summarize, the only hard data generated were the results of the composite examinations. This is not surprising in light of the college's chosen strategy, which was to lay the foundation for collegewide acceptance of a more systematic evaluation of the general education curriculum. By minimizing the emphasis on "evaluation" and putting primary emphasis on getting faculty members involved in assessment activities, the hope was that subsequent programwide evaluation efforts would reap greater long-range benefits than had they been initiated early on.

Conclusions

Policy Development and Decision Making

What have been some of the policy development and decision-making outcomes of North Adams State involvement in the APEP project? Speaking only to the program level, one concrete result was the revision in spring 1981 of the preamble and category descriptions for the general education curriculum. Because of the greater awareness on campus of the importance of higher-level thinking, the preamble makes reference to the importance of reasoning ability, and the "capstone" category of the curriculum calls for students to demonstrate the intellectual abilities of communicating, analyzing, synthesizing, and valuing.

At the time this case history was written, the curriculum committee was beginning a thorough review of the general education curriculum that would include clarifying criteria for reviewing and screening general education courses and establishing more rigorous procedures for determining which courses should be part of the curriculum. Among the criteria almost certain to be applied would be the extent to which a given course is designed to foster at least one of the intellectual abilities identified by APEP.

These developments can be fairly described as being a natural outgrowth of the college's participation in the project. When a substantial number of faculty members became involved and were challenged to think about better ways to foster higher level thinking in their courses, it became a logical extension of their activities to institutionalize a more explicit set of policies having as a common goal better communicating, analyzing, synthesizing, quantifying, and valuing.

Of course, much remains to be done. It is one thing to establish policies, quite another to implement them. Given a solid core of faculty support and the experience gained during the life of the project, successful implementation certainly is more probable than would otherwise have been the case.

Some Reflections and Recommendations

Thinking back on the activities, strategies, and decisions made over the 2-1/2 years of the project, several reflections and recommendations come to mind. At the beginning of the project, it might have been best for the college to proceed with the original plan to have groups of faculty members seek to define more clearly the various intellectual abilities implicit within the category descriptions of the general education curriculum. Very likely, these would have included those identified by APEP, but it would have been better for the faculty to "discover" this than to have them imported from outside.

The preparation of four position papers during January 1980 was a good idea as was the follow-up conference at which they were discussed. In restrospect, it would have been better to work on identifying the components of the intellectual abilities rather than stopping with the descriptive papers. The price paid was that the faculty never had full ownership of the components subsequently selected from among those developed at other institutions.

The strategy of first working with faculty members at the course level was a sound decision and constitutes the major recommendation to be made to institutions of comparable size and characteristics. Faculty members responded enthusiastically and followed through with no reward other than their own feelings of professional growth. Eventual policy decisions were an outgrowth of the entire process rather than the result of a comprehensive, programwide evaluation that might not have been so well received.

Part of the strategy developed in April 1980 called for faculty members to work in small support groups consisting of three to five members convened by a member of the steering committee. Such groups are recommended strongly, even though North Adams State was unable to implement them fully. Steering committee members simply did not have enough time to assimilate all of the project literature and to work with their groups. Should other institutions about this approach, it is recommended that they try to provide some release time for group leaders.

It is extremely important for an institution to be aware of its own faculty's needs and motivations and to use the APEP guidelines as a reference tool rather than as a recipe. The steps of the guidelines are helpful in strengthening the process of program evaluation but should not be employed without constant reference to the many forces at work within an institution.

Perhaps the greatest payoff of North Adams State's involvement in the APEP project was the contribution it made to faculty morale. Again and again, faculty members spoke of the value of meeting with others to talk about teaching and learning.

Written By

Richard E. Markham

Institutional Contact

James R. Roach, vice president for academic affairs
Richard E. Markham, chairman, Department of Interdisciplinary Studies

Ramapo College of New Jersey

Institutional Profile

Ramapo College, located twenty-five miles northwest of New York City, is one of nine state colleges is New Jersey. As it began its thirteenth year in September 1983, over one-third on its 4,500 students were part timers and more than 80 percent were commuters. Approximately 25 percent of the students were over age twenty-three.

Twenty-one major programs leading to the B.A., B.S., and B.S.W. degrees are offered by approximately 135 full-time faculty members, 90 percent of whom hold the doctorate or equivalent terminal degrees, and thirty equivalent part-time faculty members. The faculty has organized itself into six degree-certifying schools, structured nondepartmentally and oriented by academic philosophy toward interdisciplinary and theme-oriented studies: administration and business, American and international studies, contemporary arts, environmental studies, social science and human services, and theoretical and applied science. In addition, there are two academic divisions: basic studies and physical education. (As of 1985, the School of Environmental Studies was merged with the School of Theoretical and Applied Science and the Division of Physical Education was disbanded.)

Background

Although as a "second-generation" state college Ramapo had the opportunity to develop a faculty organization and culture particularly shaped by its interdisciplinary orientation, its problems and concerns were not unlike those of the other public state colleges and universities associated with the Academic Program Evaluation Project. Like most colleges over the past decade, Ramapo had endured a prolonged controversy over the nature and purpose of general education. Three elements had been especially prominent in the Ramapo discussions: the integration of critical method and subject content in disciplinary instruction; the need for demonstrated proficiency in reading, writing, and quantification; and the need for both vertical (sequential) coherence and horizontal (synchronic) coherence in courses. The decision to submit a proposal to AASCU was prompted, to a large degree, by the good match between issues prevalent in the Ramapo context and APEP's overall interests in the primary intellectual or generic competencies of communication, critical thinking, and quantification.

Other promptings, of a more external nature, occurred with the introduction in 1978 of a statewide test for all entering public college students, the *New Jersey College Basic Skills Placement Test.* This test drew public attention to "basic skills" and accelerated the development of remedial and developmental programs at all New Jersey public colleges. At Ramapo the growth of these programs increased faculty sophistication about the testing, assessment, and teaching of underprepared students. It was agreed by all that basic skills needed to be improved, but that was only the first stage of student development toward intellectual abilities commensurate with the baccalaureate.

In addition to APEP's emphasis on intellectual skills as a critical element in the baccalaureate, APEP's purposes and goals seemed, in 1979, equally timely and pertinent to Ramapo's record of institutional growth. Approaching the end of its first decade of operation, the college had experienced in its early years rapid enrollment growth and a burgeoning curriculum. The formation and later reformation of the Ramapo schools had by 1979 clarified the college's instructional content and mission. This exercise of curriculum building by a comparatively young and mostly tenured faculty was further taxed by processes that brought the majority of the faculty to tenure by 1979. During this first phase, self-study for the purposes of accreditation and certification, in addition to the cycles of a major program evaluation, also commanded the faculty's attention.

Given Ramapo's organizational and curricular growth and the great diversity of its student body, it was time to examine how the curriculum contributed to the attainment by Ramapo students of the intellectual skills of communication, critical thinking, and quantification. Given the past attention to self-study, such an examination, it was anticipated, would happily be done in concert with nine other state colleges and universities. APEP therefore became an opportunity for Ramapo to conduct a comparative examination. Viewing its own purposes and outcomes within the context of nationally representative state colleges and universities would place Ramapo's own efforts in an important perspective.

Implementation

Although faculty members and students will, at an appropriate distance, view the purposes of such a project as APEP as useful and salutatory, bringing APEP directly into the curriculum and classes was another matter. Other chapters by APEP participant institutions have made sufficient reference to examples of faculty resistance to assessment and measurement and the causes for that resistance. The universality of this situation only accentuated its importance. The challenge of how to determine the "ownership" of APEP and how to make it a legitimate institutional activity is answered largely by local governance traditions, past practices of institutional research, and the faculty culture. It soon became evident to all the APEP institutions that to avoid the suggestion of APEP as simply administrative oversight, the pathway to answering the difficult and ambiguous questions raised by APEP had to begin with and continously pass through the faculty. The APEP goals needed to be turned into intellectual questions formulated by faculty members and shaped into techniques of inquiry and measurement that would satisfy the protocols of study and research. If APEP were to have any relation to the study of teaching and learning in a faculty culture, that inquiry had to be formulated and executed by a representative faculty group.

The choice of a representative group at Ramapo followed the same lines of governance and consultation that had proven durable in the rapid development of the curriculum. Participants were invited and confirmed by the faculty of the schools and divisions. The school remained the locus of faculty governance, with all-college matters treated in standing committees comprising representative members elected by the councils of the academic units. The sturdiness of this unit governance process permitted the emergence of an eleven-person committee, which remained remarkably stable in membership during the two and more years of its activities.[1] Such committees are not simply elected. They cohere only after many hours of informal mettings, scheduled presentations at unit meetings, including the circulation of formal summary material, and passage through proven channels of faculty governance. Such a process led, at Ramapo, to the threshold of what the APEP consultants liked to call "ownership."

First-Phase APEP Efforts

By spring of 1980, one year after Ramapo formally joined the consortium of APEP institutions, the faculty committee on APEP had defined its membership and moved closer to understanding how its collective talents could be brought to bear on the intellectual and methodological issues raised by the purposes of APEP and the relationship of APEP's goals to the curricular intentions and outcomes of Ramapo's curriculum. In a summer 1980 interim report the committee summarized its initial purposes in the development of APEP and forecasted its activities for 1980-81.

- to acquaint the Ramapo community with the purposes of the project and to solicit and encourage the participation of the faculty and staff
- to review the pertinent critical literature on the definition and assessment of undergraduate college outcomes, particularly as they address the broad intellectual skill areas of communication, critical thinking, and quantification
- to examine the structure and kind of Ramapo's undergraduate instruction and to confer with the academic units about the relationship of that instruction to the project's purposes
- to propose a plan for an evaluation design to carry out the APEP goals at Ramapo during the 1980-81 academic year.

In 1979-80 the committee devoted most of its efforts to the first task of acquainting the Ramapo community with the purposes of APEP. For more than two months the committee met regularly and, with the assistance of the library staff, Ramapo's Office of Institutional Research, and consultants from within the Ramapo faculty and from the outside, pursued its inquiry into the purposes outlined above.

Definitions of Intellectual Skills

One striking phenomenon observed by the committee early on was that definitions of generic intellectual competencies are quite primitive: literature searches revealed little and the solicitation of consultants or experts resulted in few individuals who had a sophisticated knowledge of the area. The committee did find papers distributed by the AASCU Resource Center for Planned Change staff quite useful in this regard, especially Gary Woditsch's paper of generic competency and Jonathan Warren's survey of how California professors characterized their best students.

The APEP member institutions were initially invited to determine baccalaureate-level intellectual skills using the categories of communication, analysis, synthesis, and quantification. In early discussions on the nature of these intellectual abilities, committee members realized they were subject to a number of limitations of different kinds. For the purposes of this study the skills were to be exit-level rather than entry-level remedial. Ideally, they should be, in Gary Woditsch's language, "recursive and ubiquitous generic competencies." In practice, however, these abilities could not be fully abstracted from content and instructional context. The committee believed these intellectual abilities overlapped more than they were separable. In fact, it was unable to make any useful distinction between analysis and synthesis and decided that the two would best be described as "critical thinking."

Furthermore, abilities could not be merely operational or mechanical; critical abilities such as the recognition of problems were essential elements. These abilities were not to be tied into any major; they were to be generic. It was recognized that the instructional intentions of each school might emphasize different aspects of these abilities. For example, the School of Contemporary Arts might emphasize proportion and other less numerical relations in quantification than would, let us say, the School of Theoretical and Applied Science. The committee decided that its inquiry should be concerned with the student's capacity to demonstrate these intellectual abilities, not merely whether he or she had the opportunity to learn them. Although practical problems might arise, the definitions of these abilities should not, at least initially, be limited by what the faculty could teach. Finally, the committee acknowledged that its wish to define would be constrained by the economic and time limits of what is testable.

For a time the committee thought it best to talk about a constellation of elements in these skills, since the intersections of the three (communication, quantification, and critical thinking) were more common than the unique features belonging to any one of them. Critical thinking plays an essential role in the processes of quantification, and in communication one must communicate, at least to oneself, the self correctives of logic. Modeling in quantification draws extensively from the logic of analogy. Probability and statistics are frequently displayed as visual communication. Perhaps, the committee speculated, it is only the more operational features of any of the competencies that one might consider exclusive to each: computer utilization and calculation in quantification, the proper use of grammar in communication.

Baccalaureate-Level Generic Competencies

Within these prescriptions, the planning committee devised the following descriptions:

- *Communication.* This generic skill is something more than a giving and receiving of messages; it includes a number of elements such as the capacity to listen and speak well, abilities that presently exceed practical limitations of campuswide testing. The committee recognized that communication as a generic ability includes writing, reading, and visual and historical communication. Writing includes an operational knowledge of fundamentals of composition and the grammar of standard American English, sufficient vocabulary, capacity to make a thesis structurally evident in prose, and consistency of presentation. In reading, one must be able to identify the theme of a written piece of material and to distinguish between fact and opinion, literate and intentional meanings, and the connotation and denotation of terms. One must, furthermore, be able to recognize symbols and have a sufficient expressive and analytical vocabulary. Visual communication consists of the interpretation of the formal structures of visual presentations, such as graphs or works or art, and the recognition that these structures emerge from the creative interplay between materials and concept. Historical communication is characterized by its application of models taken from the sciences (e.g., forces) to historical phenomena and its elucidation of the historical conditioning of the meanings of terms.

- *Critical Thinking.* Analysis and synthesis, as has been indicated, were combined as critical thinking. As essential elements of critical thinking, problem detection, argument analysis, premise identification, judgment of logical validity, and detection of alternate premises proved more testable than critical attitudes, position taking, recognition of systematic and extrasystematic limitations on methods of thought, the use of models, and elaboration of context.

- *Quantification.* The committee recognized that the competency of quantification is a process that includes more than mathematics as it would be taught sequentially in school or college. Rather, it is processual: the application of mathematically related skills and concepts to various types of applied problem solving. These mathematical skills include counting and arithmetic, measurement, determination of geometric relationships, and algebraic, trigonometric, and statistical manipulations. These are applied in many ways, including modeling techniques and various problem solving work in the sciences and social sciences.

Evaluation Design

Despite the realization that generic competencies should suit the realities of assessment, the definitions of these competencies still proved in practice to be too lofty. Practical constraints of resources and time, and the limits of what the chosen tests purported to measure, made it necessary to draw even more restrictive lines around them to make student progress in these competencies more accessible. Further, the committee thought it excessively bold to attempt to draw up tests to measure achievement of such complex intellectual characteristics, especially when no useful validation of the additional test elements would have been available. Thus, the faculty committee felt constrained generally to use instruments already available commercially or through AASCU consultants and correlate test results with transcripts and other data already available. It was hoped that a few of the nonmeasurable features of the skills could be analyzed through in-depth interviewing of faculty members and students, coupled with a careful examination of core course syllabi and final examinations.

The committee, together with a social psychologist from the School of Theoretical and Applied Science, acting as consultant, designed in summer 1980 an evaluation of student progress in the competencies. The goal of the evaluation was to provide information on which specific combination of learner characteristics significantly predicts adeptness in communication, quantification, and critical thinking as measured by the testing instruments. These learner characteristics were listed as age, sex, academic year (freshmen through senior), SAT scores, total credits completed, and number of general education courses completed.

It was recognized that a longitudinal study was not possible within the time frame of APEP. Hence it was decided to carry on a cross-sectional study for the academic year 1980-81 with some hope that longitudinal follow-up might be possible within an extension of APEP. For the purposes of measuring written communication, a locally developed, thirty-minute essay test was devised, to be rated through two independent scores by faculty readers. This test format, as well as the rating procedure, was chosen because it reflected the essay format and holistic scoring procedures used in the essay component of the *New Jersey College Basic Skills Placement Test*. To measure critical thinking and quantification a number of comercially available instruments were examined. After pilot testing the Watson-Glaser Critical Thinking Appraisal test and the Sequential Test of Educational Progress (STEP) mathematics test with a subsample of students in August 1980, the committee decided to use these instruments. It was recognized that the student sample would have to consist of all volunteers, optimally 500 or more. A release form was drawn up to request permission for use of student's records and to match data on test scores with student's social security numbers.

Data from learner charcteristics would serve useful predictive functions. Moreover, it was felt that age, sex, and attitudes would provide useful profiles of the subjects; academic year, scores from the SAT, the Basic Skills Tests would permit measurement of value added or lost. Also, the intervalidation, or the lack of it, between scores and grade point average would be useful.

Measures in the Classroom

Once the measures had been selected, the committee faced the task of how to obtain the cooperation and participation of the students who would need to complete, in a reasonably consistent testing environment, the measures chosen. Should a student sample be recruited by letter of invitation, which may or may not include the promise of remuneration or other "incentives," or should they be tested during a scheduled class period with the understanding, of course, that their participation was voluntary? The committee decided to use scheduled class periods in the first week of the semester. The committee concentrated on those classes in which students from Ramapo's three selected schools would be enrolled and during enough hours of the day to ensure that adult part-time students would be included in the sample. To ensure a sufficient sample of senior students, the committee included as many senior seminar courses as possible. The decision to use the opening sessions of a class was motivated by the committee's belief that it would become increasingly difficult to obtain the cooperation of the faculty once course content had been fully committed. Student attendance is usually highest during first sessions, and it would still be early enough in the course for the students not to regard testing efforts as an intrusion on their attention.

Outcomes

Though the testing process was fraught with logistical problems, there were great rewards. Once a class was identified and the permission of the faculty obtained, an APEP committee member or another fully briefed faculty or staff member would appear at the class. The overall purposes of the APEP research project were outlined, and the students were assured that anonymity would be preserved and that the outcomes of the testing would be reported to any student who wished a follow-up interview. Massive defection did not materialize. In some instances individual students declined to participate and in one isolated (but still notorious) case an entire class got up and left, prompted by a faculty member who had orginally declared his interest in cooperating.

Generally, the committee found itself involved in animated discussions about the relationship between testing and learning and the function of educationl research. It appeared that students would give their time when they believed that the requested tests would contribute to institutional self-understanding and therefore to greater understanding of the intellectual abilities of the students. Long after the test period students and faculty members would return to the issues raised during the test periods, and APEP's purposes would enter into discussions of both an informal and formal nature. The testing periods gave APEP a pervasive status in the community.

Interpreting the Data

The two stages of APEP at Ramapo could be largely described as, first, a process of informing the academic community of the purposes of APEP and, second, a faculty-directed inquiry leading to the design and execution of an educational research study. A detailed summary of method and results was prepared and submitted by Ramapo to AASCU and FIPSE as the technical report in our final case study.[2] It would not be useful to replicate, in this context, the details of that report; however, an answer to one basic question can be given.

Does the baccalaureate program enhance student's abilities in the generic competencies of communication, critical thinking, and quantification as measured by these instruments?

The correlation coefficients indicated that a higher count of credits was associated with higher scores on the indices of cognitive skills. However, it must be immediately added that more advanced students have on the average more of the abilities necessary to survive in an academic program than those who have fewer credits. When test data were corrected for such factors, it seems that the program did not enhance (nor diminish) the "cognitive competencies" as measured. In order to isolate the variables used for correlation with the measures, a series of regression analyses were conducted on test scores in relation to credit count, age, grade-point average, sex, Ramapo school of enrollment, SAT verbal and SAT math scores.

The regression analyses suggested that in quantification higher credits, higher grade-point average, school of enrollment, and SAT math proved to be significant predictors of STEP mathematics testing scores. Age and sex were or lesser significance: "adult" students did not perform any better of this measure than did "traditional" students.

Some measurable progress in critical thinking through the baccalaureate program was discernible. This factor was isolatable in a statistically significant way, even though grade-point average and school of enrollment proved to be significant predictors, with the verbal score the more significant or the two. Age and sex, as with the STEP test, were not noteworthy predictors.

Thus educators can take some comfort in knowing that whatever is measured on the Watson-Glaser test is enhanced to some extent by exposure to college studies. But data from the test also seemed to indicate that students with greater reading speed and comprehension did better than those less able. Whether the results are measures of reading ability or thinking ability or some generic competency from which both flow is a vexing question, which indicates the difficulty in assigning cause and description to cognitive skills.

For whatever the local essay test measured in communication ability, there was no noteworthy difference between the average scores obtained by freshmen and seniors. Grade-point average and the SAT verbal score revealed themselves as significant predictors.

Comment should be made here about the dominance of the SAT scores in all of our analyses of student competencies. For the past decade higher education has moved to improve equality of access, but the results at Ramapo suggest that we have not fared as well with "equality of competence." Academic aptitude and preparedness that a student brings to college, as suggested by the SAT scores, remain telling predictors of the level of competency in generic skills at graduation. The baccalaureate experience does not appear to compensate and "equalize" the different levels of preparedness revealed at entry to college. Low achieving students, as predicted by SAT, remain low achieving students on the indices used for this study.

Conclusions

Many questions could be raised in response to the above outlined results and their interpretation. Faculty members who work with students over the span of a baccalaureate program are struck by the evident progress made by certain students who do not initially manifest academic abilities. Their sense of themselves strengthens and the maturity of their perceptions enables them to display greater analytical and expressive abilities. Such "evidence" of progress suggests again that in the measures used only a small range of abilities were identified and that, although the narrowly psychometric intentions of the measures and their interpretations have been satisfied, the constellation of variables that contribute to intellectual growth remains larger and more elusive than these measures could grasp. In the case of the local essay used to assess communication, for example, there is now the strong suspicion that the format and scoring of that essay are insufficient to address the complexity of the competencies embodied in communication. Whereas a thirty-minute essay on a general topic rated on a four-point scale may be useful for initial placement purposes, such a measure is inadequate to address the scope and range of abilities that can be manifested in writing, no less in the comprehensive competency of communication.

It was the committee's judgment at all stages of its inquiry that much remains to be done in the development of measures of college outcomes that have wide acceptance and credibility among the academic community. It may well be that cognitive competencies ascend an asymptotic curve in which enormous progress is made in one's childhood but abilities level off in later years. Is the function of the college curriculum, then, to provide more and more content for abilities already well-established? Or can colleges do more to receive a reasonable return in values added to cognitive competencies?

Policy Directions: A Curricular Model for the Future

APEP has in part contributed to a more cooperative institutional research environment in which faculty members are no longer so suspicious about assessment as an intrusion into the progress of their students and have some limited faith in psychometric techniques. More important, however, it has provided a base of data from which longitudinal studies of the college's effectiveness in the teaching of generic competencies can be assessed. Although no specific decisions have been made about the thrust and method (or--importantly--*funding*) of such a study, it is likely that the research would be more limited, concentrating on the effectiveness of selected general education courses in developing generic competencies.

Programs in writing and elementary mathematics have for some time now been well underway. It is to be hoped that continued emphasis at Ramapo on these programs will produce some noticeable improvement in its students' competencies in these areas. There has been increased attention to peer tutoring in composition and an emphasis on problem solving and mathematical models in instruction in the basic studies division. Much remains to be done: writing should be required and carefully corrected in more classes in the disciplines. Still to be devised is a satisfactory and economically manageable definition of exit-level competency in quantification.

One might immediately object that the students already engage in much critical thinking in the disciplines. And there is no doubt that they do. But it is left to chance that the skills learned in the disciplines become generic competencies: it is left up to the student's own devices to incorporate that skill at a higher, more complex level so that it becomes ubiquitous or tranferable to other areas.

A New Program

The college is considering a program that addresses thinking as such. A sequence in instruction in thinking would consist of two stages: a one-semester course at the beginning of a student's college career, and a second semester somewhere toward the middle of the student's four years. The first component concentrates on the behaviors universally present among those who do well on IQ tests, with some introduction to the rules governing correct thinking. The second component would directly address the canons themselves.

Most students who have been admitted to college have by that fact alone demonstrated a persistence and raw intelligence that can serve as the stuff for future development in such a course. The first component of the curriculum is meant to free up those abilities, to unblock those behaviors which inhibit the flow of such a capacity. Or conversely it is meant to inculcate those behaviors universally characteristic of high scorers on IQ tests: selective attention, sustained analysis, analogizing, suspension of closure, and autocensorship. Students who have already demonstrated this self-command by virtue of high scores on the SAT verbal test should be exempted, of course, because such work would be redundant. Such instruction does not address the actual forms of reasoning; the student would receive an introduction to precision in the use of language, an elementary use of the logic of necessary and sufficient conditions, some practice in elementary forms of argument, and some attention to models.

It is in the second component of the sequence that the methods of thought are addressed in their essentials: more complex uses of connective terms and quantifiers; case studies of patterns of scientific inference, including a look at the logic underlying statistical reasoning; a study of the patterns of derivation, including the use of matrices to establish the interconnections among many variables; the study of particular models, such as systems; and practice in problem solving, using such techniques as the elimination of alternatives and working backward from a known solution to its givens.

The college could evaluate the effectiveness of this program by comparing the scores in original SATs to scores from GRE tests. What steps can Ramapo take to reasonably ensure that such a sequence in thinking would noticeably improve students' scores on the SAT-GRE? Or, to put it in more fundamental terms, what are the criteria that must be satisfied to have an instructional program of high quality in thinking?

Besides the presumed criteria of attentiveness to detail of deadlines and academic rigor, other criteria should be pointed out. Course sections should be small enough for discussion and adequate enough in number for students to take the courses when they need them; course content should be integrated horizontally with other remedial or discipline-oriented courses the student is taking at that time and vertically integrated with other core courses; in addition, faculty and student tutors must have the special training necessary to assist students in knowing how to think more clearly. But the most important criterion for quality is that the courses must have the proper instructional modalities. This instruction must have at least three elements: the students vocalize the processes of their own thought in trying to solve problems presented to them; they have the opportunity to compare their thinking processes with models of good thinking, perhaps available on tape or from peer tutors, and are encouraged to take notes on the differences; and they are provided with feedback sufficiently detailed to be corrective and yet encouraging. Numerous experiments have shown these modalities to be successful in dramatically increasing students' scores on IQ and aptitude tests.[3]

Notes

[1] The faculty members of the APEP committee selected by the six schools and two academic divisions were Gordon Bear (social psychology), Eugene Beyers (psychology), Ronald Brady (philosophy), John Robert Cassidy--project director (philosophy), Robert Christoper--campus coordinator (literature), David Freund (photography), Thomas Goss (history), William Makofske (physics), David Turnage (physical education), Warner Wada (painting), Gabriella Wepner (mathematics). Helen Rogers (academic affairs) was the committee's administrative secretary.

[2] A copy of the technical report is available on request from Robert Christopher, Assistant Vice President for Academic Affairs, Ramapo College of New Jersey, Mahwah, NJ 07430.

[3] See, for example, Whimby's presentations on work by Benjamin Bloom and Lois Broder at the University of Chicago from 1945 to 1950 and his own work at Berkeley in 1966.

Written By

John Robert Cassidy and Robert Christopher

Institutional Contacts

Robert Christopher, vice president for academic affairs
John Robert Cassidy, professor of philosophy

Southern Illinois University at Edwardsville

Institutional Profile

Southern Illinois University at Edwardsville (SIUE) is one of the two four-year institutions constituting the Southern Illinois University System. The campus was established in 1949 when courses were offered in the metropolitan East St. Louis area. In the fall of 1965 major academic operations were centralized on the new Edwardsville campus. SIUE is located on 2,600 acres of rolling lands and woods flanking the Mississippi River southwest of Edwardsville. Eight buildings presently compose the main academic core.

There are approximately 560 full-time faculty members, 75 percent of whom possess doctoral degrees. Although many are distinguished nationally and internationally through special talents, publications, and research, teaching is the primary mission of the campus. Awards are made annually for outstanding teaching.

Enrollment at Edwardsville for the past eight years has annually averaged 10,000 students, with eighty-four Illinois counties, thirty states, and twenty-eight foreign nations represented. Nonetheless, SIUE is primarily a commuter campus, and most students live in communities within a 100-mile radius of the campus.

Background

Southern Illinois University at Edwardsville became interested in APEP in 1979 through the initiative of Earl Lazerson, then vice president and provost, and now president. The university administration and others knowledgeable about the project were excited about SIUE's selection as one of the ten pilot institutions by AASCU. The project afforded a unique opportunity for institutional self-examination.

During the 1978-79 academic year, a General Education Task Force was commissioned to begin the study of the general education program at SIUE. John Barker, professor of philosophy, was appointed chairperson of the task force. This group began its study of the general education program approximately six months before APEP activity began.

Since Earl Lazerson became president at about the time SIUE was notified of the APEP participation, Earl Beard, acting vice president and provost, became campus director for the project. Janet McReynolds, who had drafted the miniproposal, was selected as campus coordinator.

John Barker, Janet McReynolds, Earl Beard, and Earl Lazerson met to decide the appropriate structure for the project on the campus. It was decided that a separate committee should be formed to conduct the project. Subsequently, the project director and the campus coordinator met with the college deans to explain the project, to solicit their cooperation, and to request that they appoint a person from selected departments to constitute the committee. Initially, none of the committee members, including the campus coordinator, was given release time for the project. In retrospect, these two early decisions affected the committee both positively and negatively. The decision to keep the committee separate from other campus committees was probably sound.

The APEP committee membership on this campus was structured around departmental and school divisions and included faculty members from English, philosophy, business, speech communication, mathematics, biology, education, fine arts, sociology, and nursing. John Barker, chairperson of the General Education Task Force, agreed to serve as ex officio member of the APEP committee, as one way of eventually integrating the work of both committees. John Reiner, director of institutional research, agreed to serve on the committee to assist the committee in gathering necessary statistics. The committee also felt that a professional with a background in tests and measurement was vital to the project. To this end, the campus director attempted to recruit campus specialists who might assist the committee.

Implementation

Defining Purposes

To help direct its activities, the committee decided first to delineate AASCU purposes and SIUE purposes:

AASCU Purposes

- Assist ten institutions in defining conceptually and operationally four or five generic cognitive skills as generalizable outcomes of the baccalaureate degree

- With the assistance of ten diverse state colleges and universities, conduct a pilot project on academic program evaluation with the intent of improving the educational content and the evaluation procedures for undergraduate programs
- Advance the state of the art in undergraduate program assessment.

SIUE Purposes

- Develop faculty and student awareness of the cognitive skills as generalizable outcomes of the baccalaureate degree
- Pilot test, and perhaps recommend an assessment plan for, systematic undergraduate assessment
- Encourage faculty members to examine courses in the curriculum with respect to the suggested definitions of cognitive skills, which will be helpful in
  --discovering at what points in the university experience these skills are enhanced
  --discovering appropriateness or inappropriateness of the committee definitions
  --assisting curriculum committees to determine which skills are important to their curriculum, through both formal and informal means.

The members of the committee believed that the purposes, which should ultimately improve programmatic quality, could be achieved by

- improving awareness among members of the campus community about complex issues of systematic, programmatic assessment
- investigating currently available assessment instruments in order to ascertain appropriateness of local conceptual and operational definitions in comparison with commercial definitional assumptions
- serving as a sponsoring agency or forum to discuss issues related to undergraduate education.

The SIUE-APEP committee efforts were launched in January 1980 by the Washington project director. The stages of the model conceptualized at that time included a definition stage, a stage specifying acceptable levels of performance, a stage calling for the construction of an evaluation design, and an assessment stage. Participating institutions would independently determine the definitions and desired level of academic performance for each of the generic skills.

The project director's visits to SIUE generated several questions that subsequently appeared on other campuses involved in the project. Among them were the following: (1) Were SIUE and other participating institutions being asked to define generic skills needed or skills gained by undergraduates at the university? (2) What outcomes are expected of undergraduates? (3) Do all four-year programs enable students to demonstrate comparable generic skills? (4) To what extent should the APEP project be prescriptive, and to what extent should it be descriptive? (5) Is this project designed to be of a genuine experimental nature, and if so, why is there no control group? (6) To what extent, if any, will participating institutions be in contact one another? (7) To what extent will the results of the testing be made public, and to whom will the data be made available? (8) To whom will the testing instruments belong? (9) What will be the result of this project, and how will it contribute to the literature of assessment; or will it duplicate the literature already available?

The committee planned carefully for a day-long retreat by drafting a statement of procedures and outcome expectations for the retreat. The expected outcomes for the day included: (1) the construction of a brief sketch of the definitions for each of the five skills, (2) an understanding of the various committee members' areas of agreement and disagreement about the generic skills, and (3) an identification of specific areas for additional reading and research.

In preparation for the retreat, the campus coordinator formulated a set of issues that the committee needed to confront:

- Will the SIUE committee include three, four, or five skills in the project? Should the possible existence of other skills be considered?
- How can faculty members not on the committee be involved meaningfully in the project? To what extent should there be general faculty involvement? How should the committee release and disseminate information to the faculty? What steps can and should be taken to allay faculty fears about program evaluations?
- Should students be involved in decision making by the committee, and if so, how? How should students be involved in the assessment process?
- Can the project committee design an instrument that assesses the generic skills competence acquired in the major? What will the assessment scope be for the SIUE committee?
- Should the project committee inform the university of its findings? If so, by what means? Should the project committee attempt to influence the course of events at SIUE? If so, how?

According to the paradigm, the stages for the APEP project included the following: (1) faculty definition of desired intellectual abilities, (2) faculty definition of performance for each ability, (3) identification of where and how the abilities are taught, and (4) administration of multimeasure assessment instruments.

Formulation of the Definitions

The retreat proved profitable for the committee. The work of one committee member raised many questions about the cognitive skill of synthesis. Does creativity need to be defined? Does synthesis occur on different levels? Are analysis and synthesis really separable? Is there one skill that encompasses both synthesis and analysis? How would that one skill differ from evaluation? Can the generic skills be defined conceptually without considering the operational definition? To what extent do the attitude and educational philosophy on the instructor lead to the development of generic skills? An early decision made by the committee was to combine analysis and synthesis. The skill combining these processes was termed "critical thinking and reasoning." The Southern Illinois University at Edwardsville committee also decided that each of the generic skills definitions would be viewed as having two parts: the intellectual concept and the operational or behavioral definition.

Stage I was the development of conceptual and operational definitions for the generic skills. One of the generic skills, communication, generated vigorous debate. Some committee members contended that if communication were defined too broadly, its meaning would be weakened. Additionally, many questions about communication arose. For example, what kinds of organized systems can be considered communication? Does communication involve the intent of the speaker to ask an audience to make an inference or receive an experience? Does communication require that two persons or groups be consciously and overtly engaged in transmitting and receiving information?

Much of the work of creating two-part definitional statements for the skills was done through subcommittees that included members from several disciplines. This sharing of ideas with colleagues from other disciplines and with the university as well as colleges from other institutions provided opportunities for professional growth and development.

As the subcommittees developed their ideas, it became clear that SIUE definitions would be comprehensive. There was no attempt at this stage to determine those cognitive skills needed by students or the performance levels for generic skills competence. However, the Southern Iillinois University at Edwardsville definitions did attempt to include the widest possible range of skills.

Soliciting Comment and Support

After the local definitions had been formulated, revised, and carefully reviewed by the committee, the definitions were forwarded to all faculty members in the university, with instructions to return comments to the committee. Approximately ten faculty comments were received. Additionally, the definitions were published in a campuswide newsletter in an attempt to stimulate dialogue on the generic skills. A call for position papers was issued on the same topic; five were received.

Because several of the AASCU/APEP committee members were also on the contemporaneous General Education Task Force, and because the task force had created a subcommittee that also was defining the skills necessary for students in the general education program, there was considerable debate among task force members about necessary skills. The skill debated most was communication, with the debate reflecting the two viewpoints prevailing on campus: communication is a formalized presentation of opinions and values, and communication is the two-way development of interpersonal relationships. Some compromises were necessary; the general education definition of communication was more closely aligned to the presentational formal exchange, whereas the AASCU/APEP definition was more closely aligned to the interpersonal view.

In retrospect, it appears that if the impetus continues beyond the project's termination date, there should be a renewed effort to gain broader acceptance of the generic skill definitions. Additionally, there should be a reworking of the generic skill definitions in order to create an orderly hierarchy and a clear priority for assessing the skill subcomponents.

The SIUE AASCU/APEP committee spent one full quarter-year on the determination of the assessment mission, the selection of scope, and the design of the assessment plan. Each decision at this stage was related to several other decisions; consequently, the process was reexamined several times. Because this portion of the paradigm was emerging and the campus directors and coordinators were both learners and implementers, this stage was beset by the greatest uncertainty.

As mentioned previously, it was in Stage II that the need for a hierarchical structure for the skill definitions became apparent. Inasmuch as the SIUE definitions were comprehensive and global, it was clearly impossible to select all aspects of all skills for assessment. Consequently, it was necessary to establish priorities for assessment of the skills.

Each aspect of the decision-making process in Stage II was constrained by time. The decision to create assessment instruments or to use ones already available was made (primarily) on the basis of the time constraints of the SIUE faculty--none had release time for the project--as well as on the basis of the project completion date. Additionally, as the scope of the project broadened from general education to the total undergraduate experience, the complexities of Stage II broadened, presumably because generic skills are learned throughout the entire undergraduate experience.

The SIUE committee decided on the following assessment and program portrayal plan:

- A random sample of 150 freshmen and 150 seniors was generated by computer. The freshmen sample pool consisted of entering students admitted by September 1, 1981. The senior sample pool consisted of students who had completed 144 quarter hours by July 1, 1981.
- Students from various general education courses were assessed. Course selection was based on enrollment, a two-hour time block for the class, and the cooperation of the instructor. Students were assessed in the following general education classes: Oral Communication of Ideas, College Algebra, Freshman Composition, Critical Thinking, Introduction to Art, Statistics, and Introduction to Sociology.
- All faculty members were mailed a questionnaire entitled "Pilot Project of the American Association of State Colleges and Universities." The questionnaire was accompanied by a cover letter from the vice president and provost explaining that the purpose for the questionnaire was to gather program information but not to "evaluate" or compare programs with one another.

Each student received a letter from the president of the university requesting participation in the study and explaining its nature and purpose. After this mailing, the project coordinator sent a form letter to the students announcing test dates and explaining how to sign up for the testing. Testing was conducted between November 1 and December 5, 1981.

The assessment instruments consisted of multiple-choice tests in quantitative thinking and critical thinking and reasoning and one open-ended essay test based on a video presentation designed to assess the skills of communication and evaluation. The portrayal questionnaire was a multiple-choice instrument that asked students a variety of questions about class time, testing in their classes, classroom activities, assigned work, and instructor emphasis on the generic skills as defined by the committee.

All subjects were selected randomly and completed the assessment instruments and the portrayal questionnaire at one sitting prearranged not to interfere with their class schedules. Subjects in the general education classes completed the assessment instruments within the class but were given the portrayal questionnaire to complete and return the following day.

The SIUE committee decided on the following research hypothesis:

> Senior students will demonstrate significantly greater ability to use generic skills than freshman students.

The corresponding null hypothesis stated:

> There will be no significant difference in demonstrated ability to use generic skills between freshmen and seniors.

The program portrayal questions formulated by the SIUE committee were:

- What perceptions are held by selected faculty regarding the opportunity for students for students to learn five generic skills in the courses they teach?
- Does any particular cognate cluster of courses differ significantly from any other cluster regarding these perceptions?
- What perceptions are held by students enrolled in courses, regarding the opportunity for students to learn the five generic skills?
- Does any particular cognate cluster of courses differ significantly from any other cluster regrading these perceptions?
- To what extent is there agreement in perception (regarding the teaching and learning of the generic skills) between the faculty and students overall and within clusters?

The SIUE AASCU/APEP committee was unwilling to establish performance levels at the end of Stages I or II. In part, this resulted from the committee's view that the pilot project was a descriptive, trial experience. Additionally, the establishment of performance levels apart from specific instruments seemed to be arbitrary. The committee believed that desired performance levels could be recommended after the trial experience.

The committee had the greatest difficulty selecting scope, skill subcomponents, and the content domain for Stage II. The design of the assessment process at a large university is also a complex consideration. The committee considered several options before finalizing its plan. Among those options were:

- to sample core courses required in all or some of the academic areas having fifty or more majors in the program (thirty disciplines had fifty or more majors in FY 79-80)

- to select a sample of twenty graduating senior's transcripts from the pool of thirty academic areas having fifty or more majors
- to choose two courses from seven of the thirty disciplines having fifty or more majors
- to administer the portrayal questionnaire to selected graduating seniors
- to sample the general education program by administering the portrayal questionnaire in fourteen general education courses having an annual enrollment of 500 or more.

The committee decided to undertake program portrayal with the belief that the task was necessary in order to have a clearer understanding of the meaning of the assessment results. Though the committee does not regret its decision to undertake program portrayal, this portion of the project proved slow and difficult. Program portrayal on the campus was obtained by two means--the college student experiences questionnaire developed by C. Robert Pace at the Laboratory for Research on Higher Education at UCLA, and an SIUE questionnaire designed to elicit attitudes and perceptions about teaching and learning generic skills.

The committee spent a considerable amount of time designing, revising, and pilot testing a combined faculty/student questionnaire for program portrayal. The program portrayal questions listed above guided the design of the portrayal instrument.

Selection of Measures to Assess Generic Skills

The SIUE committee decided in Stage I that analysis and synthesis were inseparable cognitive processes. As decribed above, critical thinking and reasoning emerged as the integrated skill description for analysis and synthesis. The committee examined alternative commercial instruments and considered devising instruments locally. However, the overwhelming magnitude of the task was such that the members were unprepared to design the instruments, and the deadlines were too short to do the task properly.

After many meetings and much deliberation, the committee decided to select analysis and synthesis questions from the ETS pilot tests with the highest reliability and test the items with several groups of SIUE students during the summer quarter of 1981. The pilot testing led to a few revisions. By fall 1981 the critical thinking and reasoning test consisted of 11 items, ten of which came directly from ETS.

After examining many tests for quantification, the subcommittee on quantification, consisting of faculty members from philosophy, business, and mathematics, decided to prepare an instrument probing these concepts: ration, percentage, compound interest, geometric figures, equations (simple, linear), set inclusion and exclusion, statistics (mean and median), approximate numbers, distance-rate-time, and probability. Each member of the subcommittee prepared at least one problem for each of the ten concepts. Eventually, twenty-eight problems representing a wide range of difficulty were selected, and the time allotment was set at thirty minutes.

After the pilot testing, it was evident that thirty minutes was inadequate. Item analysis showed that some individual problems correlated negatively with overall achievement, presumably owing to faulty wording, a misleading diagram, or the like. Some questions were revised and others omitted, resulting in a test consisting of twenty questions.

A second pilot testing occurred with the new test in a senior mathematics class with seventeen students. The scores ranged from four to thirteen correct answers with a mean of 8.6 and a standard deviation of 2.6. An item analysis revealed a positive correlation between items and overall achievement on the test.

The quantification test was the first test administered during the testing sessions. No student answered all twenty items correctly; no one answered all twenty items incorrectly. The time allotment of thirty minutes was adequate.

Because the committee was opposed to a standardized test for communication, it set about to investigate alternatives. The subcommittee chose to assess communication and evaluation together by asking studentto evaluate the effectiveness of messages in a videotaped debate. This seemed an improvement over more static and artifical writing and evaluation assessment instruments. Consequently, a video excerpt of fifteen minutes was extracted from a debate between Dr. David Rossen and Mr. Ralph Nader on nuclear energy.

As might be expected, the issue of how to evaluate the open-ended questions in two generic skills--communication and evaluation--proved to be an extremely complex problem. Students were asked five open-ended questions after viewing the video presentation. They were given the opportunity to take notes. The pilot testing of the communication/evaluation instrument took place during the summer 1980, and several questions were rewritten as a result.

Data Collection

The assessment design chosen by the committee involved the random selection of 150 reshmen and 150 seniors. As with other institutions that had used the random sampling procedure, considerable attention was given to the proper means of soliciting student cooperation. A gift package comprising campus tickets, passes, and coupons approximately equivalent to ten dollars was offered to students for participation. In addition to receiving a letter from the president of the university and the campus coordinator, students were telephoned by faculty members on the committee. In spite of all efforts, only forty-two freshmen and twenty-nine seniors actually participated from the randomly selected pool of 300 students.

The committee members had modest success in obtaining the cooperation of faculty members who teach general education courses. As a result, 267 general education students completed the testing during the class sessions. Thus, the total SIUE sample was 338 students. The faculty program portrayal instrument was sent to 536 full-time faculty members in the fall of 1981. One hundred seventy were returned.

Data Analysis

The communication/evaluation results were analyzed according to a holistic scoring procedure adapted from the College Board testing program entitled *The English Composition Test with Essay.* Two persons unrelated to the project were trained by the committee to score the student essays according to agreed criteria. Two scores were generated: a communication score and an evaluation score. The critical thinking and reasoning and the quantification tests were analyzed separately in order to answer the questions formulated by the committee. The college student experiences questionnaire was analyzed according to the subject categories in the instrument. All data were placed on computer tapes to facilitate statistical analysis.

Outcomes

Generic Skills Assessment

The SIUE/APEP committee recognized from the beginning that the most appropriate research design to probe generic skill improvement of college students would be a longitudinal study. Because difficult time and budget constraints precluded such a choice, however, compromises had to be made. The committee's choice of studying concurrent groups of randomly chosen freshmen and seniors along with certain general studies classes still provided significant, if less than surprising, clues to the effects of the college experience on SIUE students. Comparisons were matched to control for ACT scores and age. Results of interscale correlations generally supported the validity of individual scales, with the one caution that "communication" and "evaluation" scores--derived from a written instrument--may be measuring the same skill rather than distinct skills. Reliability of the scales was shown to be high.

In each skill category (critical thinking, quantification, communication, and evaluation) scores were significantly higher for the random senior group than for the random freshman group, with the largest gain shown in quantification. General studies freshmen scored somewhat higher than random freshmen on all tests except critical thinking. After controlling for age and sex, year in school continues to be positively related to test scores, but slightly so. The difference between freshman and senior level skills--although small--persisted after controlling for ACT scores. The data supports the hypothesis that students can be expected to experience moderate but consistent improvement in generic skill level while at SIUE, although analysis of the data in light of ACT scores discloses that college attendance does not result in a major equalization of skill levels.

College Student Experience Questionnaire

The data generated by the Pace instrument seemed to confirm that the student sample was demographically representative of the SIUE population and provided a close "fit" with the university's existing self-image. In short, the results were consistent with what might be expected of an urban commuter school; students reported relatively low involvement with the university other than class-related activity. However, on about half of the activity/involvement items--including library use, faculty content, course learning, clubs and organizations, personal development, and discussion of topics in conversation--seniors reported significantly higher levels of participation than freshmen. In a broad array of areas, seniors are somewhat more involved with the university than are freshmen. A revealing section of the instrument asks students for a self-report on gains in eighteen skill and appreciation categories. Students were highly positive in eight of the eighteen areas and perceived moderate gains in four others. In the context of AASCU's APEP, it is significant that some of the larger self-reported gains occurred in areas relating to communication (writing), synthesis (putting ideas together), and evaluation (understanding self and others and developing values and ethical standards).

Moderate gains were recorded in analysis (thinking critically and logically) and quantitative thinking. In virtually all academic skill areas, seniors exhibited substantially greater gains than did freshmen.

Course/Program Portrayal

The committee designed its own instrument to portray faculty and student estimates about where, when, and how well generic skills are addressed in the curriculum. The faculty, not surprisingly, considered SIUE's courses more helpful than did students in all skill areas except those related to quantification. Faculty members thought their courses were most helpful in developing skills in critical thinking and nearly as helpful in communication and evaluation subskills. They were least positive about the usefulness of their courses in quantitative learning. Students, on the other hand, ranked quantification skills as most helpful, but differences among skill areas were small for students. Student and faculty perceptions diverged on the helpfulness issue nearly across the board. Also, faculty generally saw their courses as emphasizing critical thinking more than fact recall, whereas students viewed the courses the opposite way. Finally, the data reflecting emphasis on skills were analyzed in light of faculty disciplinary groupings. Variation among emphasis patterns was found to be so large that the committee was forced to conclude that a sound general education or distribution-among-disciplines program is necessary if students are to be uniformly exposed to those skills in their educational programs at SIUE.

Primary Outcomes of the Project

1. We now know more about the skills profile of SIUE students. The college experience does seem to make a difference, and the institution appears to be fulfilling its mission to the community and, most important, its responsibility to the students who place their trust in the university experience. Still, some qualifications and reservations are necessary. We don't know precisely, for example how much change is, attributable to normal maturation processess, how much to different abilities of current freshmen compared with freshmen of three to four years ago, and how much to the simple fact that grading and academic suspension practices of the university "select out" less skilled students, presumably leaving a more skilled senior class from which to select subjects.

2. We now know how valuable it is for faculty members from diverse disciplines to engage in a continuing, basic dialogue about what a college education means. Much controversy and more than a little impatience were evident as we discovered we often differed widely on what students should know and be able to do. Perhaps more frustrating, we found ourselves in frequent disagreement about pedagogical values, styles, and philosophies. Faculty members perhaps assume too much about campus consistency on such issues when not faced directly with empirical evidence of diversity. In this group, frustrations and conflicts were usally translated into constructive decisions.

3. We now know how difficult it is to get faculty members to sustain commitment to a long-term project. This is especially true when the service is voluntary, involving frequent meetings, long hours, constant decisions, impending deadlines, and written reports.

4. We now better realize the importance of a consistent master paradigm for a project of this magnitude. On several occasions, the local APEP committee was unsure whether the original paradigm was still in force or a newer version was being implemented. For example, at one point the description of project intent seemed to change from a descriptive pilot study ("What does basic skill learning look like on our campus?") to a diagnostic/evaluative study ("How can these results be applied to suggest program modifications in the institution?").

5. We now know from program portrayal instruments much more than we did about the perceptions of faculty members and students concerning how and where the generic skills are addressed in classes. Further analysis of the data should yield interesting comparative conclusions; for instance, do students and faculty members in the same course (or program, or major) have similar perceptions of which skills are stressed? Do faculty members believe evaluation skills are learned in a class when students have not noticed their being emphasized or taught?

6. We now know much more about the development of "home-grown" instruments and their integration with standardized instruments into a coherent research design. Still, we must recognize how difficult this is. The desire to be creative and original led us to attempt to measure communication and evaluation skills by a blended-purpose instrument administered in tandem with a videotaped presentation of a recent debate on a sociopolitical issue. Trade-offs were obvious: the creativity and specificity gained were perhaps purchased at the expense of administrative complexity and somewhat ambiguous validity. Yet many standardized instruments seemed impersonal, unsuited to the particular skill definitions generated in earlier phases of the study, and inappropriate to the SIUE student population.

7. We now can assume a somewhat heightened awareness among the SIUE community of the importance of generic skills assessment. The AASCU-sponsored project attempted in several ways to publicize its mission and to generate campus dialogue on this issue. Position papers were requested, committee members solicited aid and comments from colleagues, newspaper articles described project goals, and skills definitions were published in *The Observer*, a campus newsletter. We noted that even those persistent efforts were insufficient to stimulate significant response.

Conclusions

1. The APEP concept should become an on-going concern at SIUE. The current committee believes that a university has a responsibility to be a continually self-monitoring, self-aware institution. Generic skills definition and assessment enacts that responsibility and assures our community of our integrity and willingness to seek feedback on effectiveness.

2. The university should appoint a committee to evaluate the short-term conceptual and research procedures of the APEP committee. Many decisions were made with difficult time constraints, limited budget, and ambiguous information about national project expectations. A new committee should bring an objective viewpoint to bear on these decisions. Some important questions: Which policies are worthwhile and should be maintained? Which methodological procedures were suspect? How might the decision-making process be streamlined? Should an APEP committee engage in more interpersonal "team-building" activities prior to its procedural phases? How could the assessment procedures be made more replicable?

3. The university should create a local Academic Skills Evaluation Committee as an on-going subgroup of the faculty senate's curriculum council. This committee would supervise the investigation of generic skill learning at SIUE and initiate institutional policies relevant to that learning. It might start with the current APEP models and go beyond them. Careful balance of academic specializations and research preferences should be maintained on this committee, along with consistent student representation. Appointees should have a clear realization of the work involved and state a clear intention to follow through; the existing APEP committee did not encourage such a commitment and thus experienced absences, vacillating contributions, and even resignations from some members. The projected committee should become an on-load, assigned time for faculty members, and student members should receive some form of recognition for their participation.

4. The university should mandate through the Academic Skills Evaluation Committee a systematic, longitudinal study of generic skills learning. That is, a reasonably large but manageable group of freshmen (perhaps 200) should be identified and tracked throughout their college careers. Periodic assessment of their basic skill competencies might be accomplished through strategies similar to those of APEP. Additionally, such a longitudinal study should include:

- Data on informal college experiences similar to that generated by the Pace instrument. We should stay attuned to the extracurricular and cocurricular activities chosen by students, for these also affect the development of skills. Some studies indicate the informal network of college learning is ultimately more potent than classwork in the acquisition of life skills.
- A face-to-face interview component with appropriate self-report instruments. We know remarkably little about how skills learning in college feels from the student point of view. Narrative reports, journals, or interview responses might be content-analyzed to yield rich insights to supplement the hard data with a more human perspective.

5. The university should guarantee financial support and on-going administrative commitment for the longitudinal study. To begin such a project without a clear commitment to carry it through could be worse than not starting the project at all--and would be a disservice to project participants and the university community. Although the present pilot project was more descriptive than diagnostic (we were not intending to use our findings to initiate curricular revision or recommend different educational philosophies), it is not unreasonable to expect that a four-year longitudinal study, thoroughly reasoned and coherently conducted, might result in such diagnostic recommendations.

6. The university should make an on-going, systematic commitment to create campus dialogue on skill definitions and criteria for competence. A broader acceptance of and agreement on an orderly, hierarchical generic skill scheme would be helpful, and this might be accomplished through university

sponsorship of colloquia, seminars, informal publications, and open meetings. Several questions could be meaningfully addressed: What is the central purpose of the teaching-learning process? Have Southern Illinois University at Edwardsville students mastered the generic skills? Which behavioral and attitudinal changes are reasonable to expect of the Southern Illinois University at Edwardsville students? How can Southern Illinois University at Edwardsville facilitate generic skill learning to make it both efficient and humane?

Written By

Robert Anderson, John Farley, and Janet McReynolds

Institutional Contacts

Robert Anderson, professor of speech communication
John Farley, professor of sociology
Martha Welch, professor of nursing
Janet McReynolds, assistant to the vice president and provost

State University College at Potsdam (NY)

Institutional Profile

Potsdam College was founded in 1816 as a small academy that later became a Normal School in 1867. In the 1940s, the college became a Teachers College and in 1962, a college of Arts and Science in the massive State University of New York system. It is currently a comprehensive college with its strongest programs in music, computer science, mathematics, economics and education. The undergraduate population of the college is now 3,600 students with an additional 600 part-time graduate students.

Background

At the time that APEP was being proposed to member colleges of AASCU, administrators at Potsdam College were already concerned about intellectual skills. The college's most recent campus mission statement contained statements on the importance of student skills, and the staff had begun rewriting the mission statement and wondering whether these objectives really were fulfilled. A desire to understand better the outcomes of our educational programs through evaluation led to Potsdam's application to the AASCU Resource Center for selection as one of the ten state-supported colleges to participate in the APEP. We were particularly interested because the vice president for academic affairs and the associate vice president had both participated in the summer institute of 1978, when work began on skills development.

Implementation

In the fall of 1979, an already existing Committee on Academic Skills was abolished and a new Committee on Intellectual Skills was formed. It included key campus leaders and representatives from every significant academic area. This committee began work in October 1979 and quickly became a part of campus life.

Defining Generic Skills

The committee spent the first 2-1/2 discussing and reaching consensus on a conceptual definition of the four skill areas to be included in APEP. These were analysis, synthesis, quantification, and communication. Faculty and administrative staff members alike found this experience to be highly stimulating, because they normally do not set aside time to sit and discuss educational outcomes from the perspective of a wide variety of disciplines. The committee did conduct these discussions and, in the process, determined that valuing was an additional skill that should be studied in the project.

In February 1980, the committee began reexamining the broad conceptual definitions with an eye toward defining subskills for each of these. The committee was subdivided across the areas of (a) natural science, (b) social sciences, and (c) humanities to discuss the subskills needed under each working definition. Although reaching a consensus occasionally required difficult concessions, this second task was achieved by the end of that spring. More importantly, this fusion demonstrated to the now-reunited committee how similar the separate concerns had been and how comparable they could become when formulating unified subskills.

Moving from Conceptualization to Operationalization

Early in the summer of 1980, a smaller but very representative version of the Committee on Intellectual Skills renewed its activities. This committee began the process of converting the definitions that had been previously determined into operational statements, choosing and modifying a research design, selecting appropriate evaluation formats, and choosing appropriate but varied evaluation instruments. This process of creating operational statements from the conceptual definitions was a difficult and frustrating task. Having spent months developing statements that characterized a valid institution-specific version of the generic skills, the members felt some ambivalence about writing simplified, testable examples of their formerly comprehensive works. This stage ultimately required careful editing of the lists that had been prepared, sometimes with disagreements among committee members but with a final consensus. The committee was not completely satisfied with its observables, but it understood as it entered the actual testing cycle that without the crucial transition from conceptualization to operationalization, the project would not later measure change in predictable or meaningful terms.

Choosing a Research Design

In creating the research design, the committee adopted a model it labeled a "Multi-Cell Research Design." It was determined by the group that the best way to measure skills might be across class years and across major academic divisions of the college. These subdivisions were liberal studies natural science, liberal studies social science, liberal studies humanities, Crane School of Music, Educational Certification Program, and undeclared students. We had been discussing the possibilities of purchasing tests, when the committee decided to call the APEP consultant, Jonathan Warren from the Educational Testing Service. Warren visited us in the summer of 1980 and explained to us the instruments he was developing as well as the choices open to us for our own observations. One choice included working with him to aid his test development, using existing instruments. Warren had been working with colleges in California but said he would welcome APEP colleges and could ultimately provide us with data about our students.

Developing and Administering the Tests

We later determined that a combination of Dr. Warren's four tests and some locally developed instruments would be the best way to go. For example, two art historians developed a computer-scorable exercise for analysis/synthesis. A history professor produced a one-act play, *The Birdbath*, after which students wrote an essay concerning receptivity and sensitivity, an exercise in both communication and valuing. During the fall of 1980, we began to draw up the lists of students who could potentially participate in the testing exercise. We also developed questionnaires to collect data about these students concerning their cocurricular learning, recognizing that students can easily perfect some of these skills in cocurricular activities. In addition, we asked students which of the courses they had taken had taught them the skills we were examining. We asked faculty members for these students to rate them on the five skills and to explain how they develop courses which lead to improving these skills. In the spring of 1981, with only the fact that they were helping the college as their incentive to appear, students were assembled and spent an entire day in the testing exercises. A total of 177 students actually completed all exercises.

Outcomes

During the summer, we began the process of tabulating information, and a group of readers spent an entire day away from the campus at a retreat center reading the essays and scoring them on the basis of communication as well as valuing. By the end of that second summer, the committee then had information ready to be studied. We were then able to make comparisons among the various cells of our research design. The four tests from ETS were scored by ETS, and Warren returned to the campus to discuss results with us. The fact that Potsdam students had performed well in his tests helped strengthen our knowledge that this had, in fact, been a worthwhile undertaking. Each of Warren's visits also rekindled enthusiasm in the Potsdam committee members.

The steering committee then became interested in further development of the locally produced test instruments. The study showed us in which year and in which academic area students scored better on these skill areas.

Conclusion

We have only slightly tapped the wealth of information we collected. We need to find time and resources to continue to analyze the data and to raise significant questions about our curriculum.

Written By

A. Paul Loucks and Bruce A. Conroe

Institutional Contact

Bruce A. Conroe, vice provost
P. William Shearer, associate professor of psychology

University of Nebraska at Omaha

Institutional Profile

The University of Nebraska at Omaha (UNO) is one of three major administrative units of the University of Nebraska. The institution was founded as a private university in 1908 and became the Municipal University of Omaha in 1931 and part of the University of Nebraska in 1968.

Organizationally, the University of Nebraska at Omaha includes six collegiate units, which report to the vice chancellor for academic affairs. These colleges are Arts and Sciences, Business Administration, Continuing Studies, Education, Fine Arts, and Public Affairs and Community Service. In addition, the University of Nebraska-Lincoln's College of Engineering and Home Economics offers programs on the University of Nebraska at Omaha campus. Graduate programs are offered through the system-wide Graduate College of the University of Nebraska.

Currently, nearly eighty baccalaureate degree programs (including those in engineering and home economics) are available at UNO. Graduate degrees available on the UNO campus include the Master of Arts, Master of Science, Master of Business Administration, Master of Public Administration, Master of Social Work, Master of Professional Accounting, Master of Music, and Specialist in Education. In addition, the Department of Psychology participates in a joint Ph.D. program with the University of Nebraska-Lincoln.

The University of Nebraska at Omaha is the only comprehensive public university in a metropolitan area of approximately 600,000 people. A commuter institution having no dormitories, UNO draws virtually all its students from the Omaha metropolitan area. The student body of approximately 15,000 is almost equally divided between men and women. Minority students account for an estimated 10 percent of the student body. Approximately 75 percent of UNO students are employed on either a full- or part-time basis, and the median student's age is twenty-six. The average composite ACT scores for 1981-82 entering freshmen was 19.0.

UNO has a full-time faculty of nearly 400. Over 75 percent of the faculty members possess terminal degrees, and many are actively engaged in research and public service activities as well as instruction.

Background

Participation by UNO in APEP was presaged by developments in three key areas of campuswide significance: general education, planning, and program evaluation. First, during the 1978-79 academic year, each college reviewed its general education requirements, a process undertaken upon the initiative of University of Nebraska President Ronald W. Roskens. It should be noted that there is no campuswide, or systemwide, general education requirement at the University of Nebraska; rather, each college within the university establishes its own requirements. At UNO, the 1978-79 review largely reaffirmed existing requirements; however, it did generate widespread interest in the general education component of the curriculum and served as the catalyst for an extended study of general education requirements by the college of arts and sciences.[1]

Simultaneously, considerable faculty and administrative attention was devoted to planning. Early in 1979, the development of college-, campus-, and university-wide five-year plans was initiated. The five-year plans were to be prepared in accordance with a set of twelve planning guidelines adopted by the board of regents. A format for the UNO college and campus plans developed under the aegis of the Ad Hoc Academic Program Steering Committee included an evaluative component. The nascent Academic Program Evaluation Project was viewed as a promising approach to satisfying this evaluation requirement, and reference was made in the draft campus plan to the possibility of UNO's participation in the proposed, ten-member APEP consortium. This planning effort eventually was superseded by a more centralized process; nevertheless, interest had been stimulated in APEP as an evaluative tool.

Finally, the newly appointed vice chancellor for academic affairs, Otto Bauer, initiated discussions in July 1979 about the formation of an academic planning council, to be charged with conducting an on-going qualitative and quantitative review of all of UNO's academic programs. Whereas this review process was conceived as focusing on departmental units and their majors, APEP was seen as offering another, complementary dimension for analyzing the undergraduate curriculum by concentrating on generic intellectual skills.

In brief, APEP emerged at the University of Nebraska at Omaha when UNO was addressing general education, planning, and program evaluation. The timing of the project, as well as its substance, coincided agreeably with campus initiatives in these important areas.

Implementation

In launching APEP at UNO, the initial task was to select the faculty committee that would guide the project to completion. Essentially two alternatives were considered--appointing a new faculty committee to assume responsibility for the project or adding APEP to the charge of an existing body. After a review of the alternatives and their implications, it was decided to invite the Educational Policy Advisory Committee (EPAC) to serve as the APEP faculty committee.

At the time APEP was initiated, EPAC had served as UNO's campuswide curriculum committee for nearly six years. The committee consisted of one faculty member from each of the colleges of arts and sciences, business administration, education, fine arts, and public affairs and community service and from the library; one undergraduate student; and one graduate student. Faculty members served overlapping terms of three years, whereas student members served one-year, renewable terms. Two faculty members rotated off the committee each year. The committee reviewed new courses and programs and advised the vice chancellor for academic affairs about a wide range of academic policy matters. Meetings were held twice a month and chaired by the vice chancellor.

Although a separate committee would have been able to focus exclusively on APEP, some concerns had been expressed about the possible proliferation of committees. Given its involvement in curricular matters, its representative composition, and its internal rapport, EPAC was the logical choice. After reviewing APEP materials, the members unanimously agreed to undertake the project.

In 1979-80, faculty members of EPAC included Dr. William Callahan, Associate Professor, Special Education; Dr. John Kasher, Professor of Physics; Dr. Janet Rives, Associate Professor of Economics; Mr. Tom Tollman, Assistant Professor and Reference Librarian; Dr. Sam Walker, Associate Professor of Criminal Justice; and Mr. Bob Welk, Assistant Professor of Dramatic Arts. The following year Professors Callahan and Welk completed their terms on the committee. They were replaced by Dr. Robert O'Reilly, Professor of Educational Administration; and Dr. Warren Prince, Professor of Music. In 1981-82, Professors Kasher and Rives were replaced by Dr. Bruce Baker, Professor of English, and Dr. Richard Ortman, Associate Professor of Accounting. Dr. Deanna Finkler, Associate Professor of Psychology, served as consultant.

The committee first addressed the scope of the project (i.e., whether the focus should be on general education or on the general outcomes of a college education). The prevailing view was that the five APEP skills, although closely associated with general education, were by no means limited to general education. Had the project been confined to general education, a complicating factor would have been the lack of uniformity among the colleges in their general education requirements. This was not, however, the reason for extending the scope of the project beyond general education. Rather, it was believed that the skills were developed at all levels of the curriculum and that this approach would afford an analysis of the effectiveness of the entire curriculum in the development of the five key intellectual skills.

The broad goals of the project also were identified at an early date. Indeed, agreement on project goals proceeded hand in hand with the selection of the scope of the project. The two basic goals established for the project were (1) to discover the level of proficiency in the five APEP intellectual skills displayed by students nearing completion of a baccalaureate degree and (2) to determine whether graduating students in the respective colleges demonstrated different profiles of the generic skills. The profiles were to be made available to colleges for their use in considering possible curricular changes. It was recognized, however, that primary responsibility for initiating curricular change rested with the faculties of the respective colleges. Moreover, given the differing curricular emphases among the various colleges, as well as the absence of a campuswide general education component, some differences in the profiles among graduates of the several colleges were expected. Finally, it was emphasized to the faculty that APEP was not intended to serve as a review of individual programs nor as a tool for reallocating resources within the institution. Although this evaluation approach could be adapted to those purposes, University of Nebraska at Omaha's goals were more modest. Essentially, it was interested in developing a diagnostic approach; the prescriptive remedies, to the extent they might prove necessary, were to be outside the scope of APEP as it was conceived at UNO.

Defining the Generic Skills

Having charted the broad goals of the projet, EPAC turned to developing campus-specific definitions of the generic skills. Discussions centered first on communication, quantification, analysis, and synthesis; the committee agreed to address valuing after completing its work on the other four skills.

To stimulate discussions, several handouts were distributed to committee members. Documents included the April 1979 issue of *Alternatives*, which contained several articles on intellectual skills; draft definitions of intellectual abilities, excerpted from a case study used in Fall 1979; regional seminars conducted by the Resource Center for Planned Change; and a summary of key elements of Bloom's taxonomy of educational objectives.

Finkler also prepared matrices relating skills, content areas, and modalities to Bloom's academic skill levels. Subsequently, additional materials on Bloom's taxonomy were provided, along with a description of Kohlberg's model of the development of moral judgment and a copy of a sample student's responses on the Rest Defining Issues Test.

After considering several ways of defining the skills (e.g., global statement or an exhaustive list of all components of each skill), it was decided to adopt Jonathan Warren's definitions of communication, synthesizing, and analytical skills as a point of departure. These definitions had been published in the April 1979 issue of *Alternatives*. Warren's approach seemed particularly well-suited to the project, in that instruments being developed by the Educational Testing Service would be based on those definitions. Moreover, since they were stated in terms of three qualitatively distinct skills (low, medium, and high), Warren's definitions effectively combined two steps of the APEP paradign (writing definitions and establishing skill levels) into one. Most importantly, Warren's definitions were viewed as clear, thorough, illuminating, and well-reasoned. Some consideration was given to abbreviating these definitions; however, the committee decided to adopt the published versions.

Attention then turned to developing a definition of quantification following the three-level approach of Warren's definitions of communication, synthesizing, and analytical skills. Abstract, curriculum-based, and tool-based draft definitions of quantification were developed. Discussions centered on the abstract and curriculum-based definitions; elements of the two were synthesized into the definition adopted by the committee.

Valuing was the last skill to be considered. Dr. William Blizek, Associate Professor of Philosophy and a recognized campus authority on valuing, provided helpful insights during the initial discussion of this skill. A number of key ideas relevant to values and valuing were considered. These included intellectual courage, intellectual openness, intellectual tolerance, sensitivity to other persons and their sensitivities, social responsibility, intellectual honesty, searching for truth, and critical (i.e., discriminating) thinking about values. Although committee members approached valuing from different perspectives, they concurred that it is a vital outcome of the undergraduate curriculum and should be addressed in a project such as APEP. A three-level definition centering on the concept of critical thinking about value-laden issues was adopted.

Upon reflection, it was agreed that a succinct definition should be developed for each of the five skills in addition to the previously adopted statements describing the three levels of performance. This brief task, which lasted approximately four months, concluded the definition-writing stage of the project. In addition, members of the committee arranged for pilot-testing in their classes of measures being developed in conjunction with this project by the Educational Testing Service.

Designing Testing Plans

During the summer, a tentative plan for the testing phase of the project was developed in some detail. This was accomplished by the project consultant, who worked the equivalent of one month for this purpose. To obtain additional guidance, Finkler arranged to meet with a faculty member of each of the colleges whose students were to take the APEP test. The faculty representatives, designated by the deans, typically were members of their college's curriculum committees. This approach was intended to identify specific college purposes for the project, in addition to the more general ones adopted by the committee, and to elicit problems peculiar to the students of individual colleges. These contacts also provided some helpful feedback regarding faculty concerns and perceptions of the project.

Two meetings were held with each college representative. During the first, Finkler described the project in some detail and posed a series of questions for further discussion. During the second session, these questions were considered and the representative's counsel sought. Among the issues discussed were the adequacy of the generic skill definitions from the college's perspective, the preferred measurement strategy, the purposes the study might serve for the college, and the college's preference in the relative emphasis on program portrayal and student outcome assessment.

The tentative plan was organized aroung three of the five charts made available at the June APEP meeting. The charts proved particularly helpful in articulating relationships between project purposes, the relevant student populations, and student sampling procedures. In addition, the project consultant prepared extensive notes on the then-current ETS measures of quantification, communication, analysis, and synthesis. The tentative plan was available at the beginning of the new academic year for review and discussion by EPAC.

Selecting Testing Instruments

During the fall semester of the 1980-81 academic year, EPAC devoted considerable attention to reviewing testing instruments (particularly the ETS measures) and considering sampling procedures and other issues raised in the tentative plan. During the committee's review of measurement instruments, the results of the pilot tests of the ETS measures became available, together with a set of revised ETS scales that reflected the results of the recently completed pilot testing.

Working individually or in pairs, committee members examined the ETS scales in depth. After extensive deliberation, the committee agreed on the following measures:

Measure	*ETS Form*	*Items*
Analysis	31	6, 9
	32	3, 6, 8
	33	1, 5, 7, 9
Quantification	27	1-9, 11-14
	21	22
	22	17
	23	20
	24	7, 20
Synthesis	44	1-12 (all)
	41	10
Communication	14	1-12 (all)

These four measures, consisting of fifty-two multiple-choice items, constituted the first part of the UNO APEP instrument.

The committee determined that, to accomplish its purposes, the test for communication skills should be supplemented by a writing exercise. After considering several possibilities, the committee decided to adopt an essay test developed in consultation with the Department of English and based on its diagnostic writing examination. This approach was particularly appealing because the department has, in recent years, developed a highly refined method of scoring essays for the purpose of placement into freshman composition courses as well as determining whether certain students meet the writing proficiency requirements for graduation. This alternative had the additional advantage of being linked to an established curricular requirement. The essay test consisted of five questions on relatively current social or political issues and were considered to represent a medium level of difficulty. Students were to select one of the five questions.

For valuing, the committee decided to adopt the Defining Issues Test (DIT) developed by James Rest. It was generally agreed that this was the best alternative readily obtainable; the availability of regional norms added to the appeal of the DIT.

Designing Sampling Procedures

The committee next addressed the critical issue of sampling procedures. A key question was whether the test should be administered during class time or given outside class during specially arranged testing periods. The obvious advantage of classroom testing was that the desired sample size could be ensured; there would be a more or less captive audience, assuming the cooperation of the respective faculty members and students. The main disadvantage was that nearly a full week of class time would be required to administer the test. Consultations with a variety of faculty members suggested this alternative was not feasible.

Administering the test outside class posed a different set of considerations. By inviting specific students, as contrasted with classes, a more exact random sample could be developed. However, the rate of response to an invitation to take a two-hour test typically is not high. Hence, a specific number of examinees could not be ensured, and an excessively low turnout could impair the project seriously. Nevertheless, it was agreed to adopt this approach, despite its inherent difficulties, with the understanding that special efforts would be made to maximize the rate of response.

The committee agreed to a sample size of 100: fifty senior students (defined as those having ninty semester hours or more) from each college, plus fifty freshmen for comparative purposes. These were to be stratified random samples, drawn to reflect differences in age, sex, race, grade-point average, and major. The office of institutional research assisted in drawing samples.

Administering the Test

Several steps were taken to ensure a response rate that would contribute to the success of the project. First, arrangements were made to administer the test at four different times during the week of April 6: Wednesday evening, Thursday evening, Saturday morning, and Saturday afternoon. It was hoped that interested students could fit one of these times into their schedules. To encourage student participation, it was decided to charge a fee of $12.50 for taking the test. Clearly, this proved to be a positive inducement, based on informal student reactions. Funding for this purpose was provided by the FIPSE-AASCU grant.

Invitations were mailed to the 300 students identified in the stratified random sample (fifty from five colleges, plus fifty freshmen). To assist in standard-setting, the deans were asked to identify students they would expect to perform at high and medium levels in each of the five skills. Approximately forty students were identified and invited to take the test, in addition to those randomly selected.

The response rate to the first mailing was nearly 40 percent, with about 80 percent of the replies being favorable. A second letter was sent to the students who had not responded to the first invitation within one week. Both letters were written in an informal style. Supplementing the mailings, EPAC members telephoned students who had not responded within eighteen days of the first letter. To provide positive reinforcement, a confirming letter was mailed to students who had indicated they would take the test. Ultimately 160 of the 340 invitees took the test, a participation rate of 47 percent. Approximately 10 percent of the students who responded favorably to an invitation failed to appear.

The test was administered in a large lecture hall in a familiar classroom-office building by a member of the project staff, who provided a general explanation of the project and answered questions at the beginning of the exam period. The staff member also read aloud the instructions appearing at the beginning of each test booklet. The four ETS measures were given first. Students were allocated fifteen minutes to complete each part of the test. After a ten-minute break, students were allocated one hour to complete the DIT and essay segments of the exam.

One of the examinees was a reporter for the student newspaper who, shortly after taking the test, wrote a major feature story about the test and the project. Occupying one-fourth of the front page of the April 15 edition, the article provided excellent campuswide publicity of APEP.

Setting Performance Standards

The establishment of performance standards on the APEP test was accomplished in two ways. First, EPAC members undertook individually to determine appropriate ranges for high, medium, and low scores on each of the ETS tests. Subsequent committee discussion revealed consensus among the members about the scores for each level of performance. Accordingly, the committee readily reached agreement of the following score ranges for the ETS tests:

Range (No. of Correct Responses)

Skill	**No Items**	**High**	**Middle**	**Low**
Analysis	9	7-9	5-6	0-4
Quantification	18	14-18	9-13	0-8
Synthesis	13	10-13	6-9	0-5
Communication	12	9-12	6-8	0-5

Existing English department standards (A=90-100; B=80-89; C=70-79; Not Passing=0-69) for the essay diagnostic test were adjusted to compensate for what was considered insufficient time for this portion of the test. The standards ultimately employed are as follows: A=85-100; B=75-84; C=65-74; D=50-64; not passing=0-49. No standards were established for the DIT.

Next, each college was asked to indicate its expectations for performance on the test by its senior students. Typically, this task was accomplished by the college's educational policy committee. Each college indicated the percent of its senior students expected to score in the high, middle, and low ranges for each of the ETS tests, and for "A," "B," "C," and "Not Passing" scores in the essay test. While colleges were to provide copies of the tests, they were asked to state their expectations in the absence of information about their students' performances on the measures in order to avoid biasing their judgments. One college's expectations are shown below as an example of how this task was approached:

Anticipated Performance (Percent)

Skill	High	Middle	Low	
Analysis	48.3	46.7	5.0	
Quantification	35.0	56.0	9.0	
Synthesis	45.0	48.0	7.0	
Communication	56.0	41.0	3.0	
	A	B	C	Not Passing
Essay	17.0	45.0	38.0	0

Analyzing the Data

The test data were analyzed during the summer by the project consultant, who prepared an extensive technical report on the measurement procedures and results. Some of the more salient findings of the report are summarized here.

Two methods of analyzing the student test scores were used. The first method compared proportions of students achieving scores in the high, middle, and low proficiency ranges identified by EPAC to distributional profiles established by the collegiate educational policy committees as representing both "desirable" and "reasonable" levels of achievement for their senior students. Chi-squared tests were used in this analysis.

The analysis revealed that, in general, the student test scores differed from the expectations of the faculty committees. Scores on the four ETS measures were both lower and higher than faculty expectations, depending on the college. In three of the five colleges, for example, scores on the quantification test were higher than expected. In two colleges, students scored higher than the faculty expected on at least three of the four ETS measures. However, it should be noted that these results were as much a function of faculty expectations--which vary across the colleges--as of student performance. In contrast to the ETS measures, student scores on the essay test were consistently lower than faculty expectations. By this method of analysis, seniors scored higher than freshmen on the quantification, synthesis, and essay tests.

The second method of data analysis addressed differences: (1) among seniors in the five colleges and (2) between seniors as a group and freshmen. For each comparison, a multivariate analysis of variance was performed. The results of these analyses indicated that only the essay and DIT "P" scores differed significantly among seniors in different colleges, and that only the essay scores differed between seniors and freshmen.

The analyses showed that seniors in the colleges of arts and sciences and business administration scored higher than those of the other colleges on the essay, and that arts and sciences seniors, scored highest among the colleges on the DIT. Seniors scored significantly higher than freshmen on the essay; however, neither group's mean score represented a passing grade of "A," "B," or "C." Although the means on the ETS measures did not discriminate among seniors in different colleges or between seniors and freshmen in this method of analysis, the means were predominantly the middle to high ranges as identified by EPAC. Although standards were not adopted for the DIT by EPAC, the analysis revealed that seniors and freshmen scored in the midrange identified by Rest (1979) as typical of college students and general adult groups.

Interpretation of the results is complicated by questions of validity of the ETS measures and by external factors, which appear to have significantly depressed the essay scores. These are addressed in detail in the technical report. Suffice it to say, it was concluded that the ETS tests did not contain a sufficient number of items to distinguish general academic aptitude from the specific generic intellectual skills. On the essay test, it appears that a longer period of time may have been needed for many students to do justice to the writing assignment. Also, informal interviews indicated that a number of students concentrated their efforts on organization and style and did not proofread their papers to correct misspelled words and grammatical mistakes, despite instructions to do so. Such errors significantly lowered the essay test scores.

Conclusions

Finkler's technical report was reviewed by EPAC early in the fall semester. Discussion addressed the possibility of conducting follow-up studies of writing skills. The prevailing view that emerged, however, was that such studies likely would shed little light on student writing skills. Although the APEP essay scores were somewhat disappointing, they were not altogether unexpected. And they did indicate significant improvement in seniors as compared with freshmen. Recognizing that responsibility for curricular change rests with the respective college faculties, EPAC decided to refer the report to the college educational policy committees for their review. EPAC also recommended that the colleges consider establishing additional writing requirements for upper-division students for the purpose of maintaining the writing proficiency developed in the lower-division English composition program.

Thus APEP formally concluded at UNO. The project produced the first comprehensive assessment of the academic skills of graduating students since the institution became a part of the University of Nebraska. Profiles of student abilities were developed for each college, and these have been shared with the respective colleges. Hence, the two basic goals of the project were accomplished. In retrospect, a considerably longer test would have produced a richer and more reliable set of data. However, a solid, instructive beginning was made and can serve as the basis for further efforts.

Notes

[1] In 1980-81, the college of arts and sciences adopted a series of modifications in its general education requirements. In addition, an ad hoc Task Force on University Requirements was appointed to consider a variety of issues, including the appropriateness of adopting a campuswide set of graduation requirements. The task force developed such a proposal; however, only three of the six colleges offering undergraduate programs had approved it as of June 1983.

Written By

John T. Farr

Institutional Contacts

Otto Bauer, vice chancellor for academic affairs
John T. Farr, assistant vice chancellor for academic affairs

Wayne State College (NE)

Institutional Profile

Wayne State College, located in Wayne, Nebraska, a city of about 5,500, WAS established in 1909 when the state legislature authorized the purchase of the property, buildings, and equipment of the Nebraska Normal College, a private institution founded by Professor James M. Pile in 1891 and operated under his direction until his death in 1909. It opened as a State Normal School in September 1910.

In 1921, by act of the state legislature, the school became a State Normal School and Teachers College with legal authority to grant baccalaureate degrees in education. In 1949, the legislature changed the name of the institution to Nebraska State Teachers College at Wayne and granted authority to confer the baccalaureate degree for study in liberal arts. The graduate program leading to a master's degree was authorized in 1955, and in 1963 the legislature changed the name of the college to Wayne State College.

Wayne State College has an average enrollment of about 2,800 students on the Wayne campus during the regular session. Summer enrollment totals about 1,000 in the May, June, and July sessions.

Wayne State College is authorized to grant the following degrees: Bachelor of Arts in Education, Bachelor of Fine Arts in Education, Bachelor of Science in Education, Bachelor of Arts, Bachelor of Science, Master of Arts in Education, Master of Science in Education, and Education Specialist in School Administration.

Background

The phrase "global village," coined by Marshall MacLuhan more than a dozen years ago, emphasized the impact of media and technology on man's global perspectives and reinforced the perception of an escalating interdependence of mankind. As a result of almost-daily technological advances, today's students are moving into an enigmatic world of values confrontation, depersonalization, interpersonal communications anomalies, conformity, and ambivalence. Students' ability to put conflicting evidence into perspective, to sort out the meaningful from the trivial, and to discriminate between the enduring and the ephemeral will be crucial to their contribution toward helping to solve rather than becoming part of society's problems.

Institutions, through their general education programs, can promote skills to be used as lifelong resources by students as they are forced to come to grips with the problems manifested by an ever-changing, complex society. These skills (or "student outcomes," as they are sometimes called) may, by consensus, be viewed as institutional goals. Several issues are unavoidable and must be confronted by each institution's faculty: (1) What is the effect of our general education programs upon our students' lives? (2) How might this effectiveness be evaluated? (3) How might principled, nonthreatening, and academically defensible changes be promoted?

Although most in higher education agree that general education programs are institution-wide programs shared by the full faculty, many faculty members have come to feel disenfranchised from the general education decision-making process. This feeling of being "on the outside" may precipitate the factionalization of an institution's faculty, "turf protecting," and obstructionism. At the same time, those who feel "on the inside" of the general education decision-making scene sometimes tend to become intransigent whenever they perceive a threat to their general education domains.

In an attempt to confront the issue of assessing student outcomes in general education, several projects have been generated. The purposes of this report are to examine Wayne State College's role in the Academic Program Evaluation Program (APEP) and to suggest a process for arriving at institutional agreement about the goals of general education, to examine the process of generating conceptual definitions in key general education skill areas, to propose a process for operationalizing these conceptual definitions, and to describe procedures for gathering longitudinal general education data that may serve as a basis for curricular generalizations.

Implementation

Establishing Institutional Goals

There is disagreement about what constitutes evidence explaining changes in attitudes, values, personal traits, and other characteristics that affect a student's college experience. That students do change as a result of their college experience is accepted. Establishing the reasons, by using an appropriate assessment instrument, is clearly another matter.

More than twenty years ago, Paul Dressel indicated there were at least two distinct approaches to evaluating general education programs.[1] One was to determine the skills of students at matriculation and graduation. Improvement in the respective skill areas under examination is thus ascribed to the particular program designed to satisfy those specific objectives. The obvious difficulty with this method is that the observed changes cannot logically be credited to the specific program because of the myriad extraneous and intervening variables that may have influenced the student during the college experience.

A second and more valid approach to program evaluation is to assess student outcomes, that is, what students are able to do, as evidenced by observable and measurable criteria, at graduation. These student outcomes (also called "performance criteria") can then be evaluated relative to the degree to which program goals are being achieved. If, for example, students do not demonstrate the desired outcomes at graduation, the curriculum can then be modified to remedy the deficiencies for succeeding classes.

Student outcomes, then, are the ends or objectives of the curricular program, whereas the instructional procedures within the components of the program are the means by which these objectives are achieved. Objectives can be constructed on a course-by-course, department-by-department, or generic skills basis. However, it must be emphasized that in the case of institution-wide programs such as general education, the larger the basis of consensus among the general faculty about specific student outcomes, the greater the likelihood of achieving the program's objectives. To this end, an institution-wide committee composed of representatives from each academic component contributing to the general education program (including student representation) was constituted for purposes of determining program objectives, performance criteria, and evaluation procedures to assess program effectiveness.

Generating Generic Skills

Once the appropriate committee was constituted, it began to function as a conduit representing its various constituencies' views and providing feedback on the particular outcomes that should be manifest in the graduating student. However, the levels of hierarchy extant in an institution, compounded by a cumbersome number of groups to be represented on the committee, made it difficult to obtain consensus on conceptual and operational definitions of generic student outcomes.

The Wayne State committee used the following approach: each committee member was to ascertain from the respective academic component those broad skills that are deemed essential and should be apparent in the graduating student. The committee then proceeded toward consensus on which of the skill areas should be addressed and subsequently, in a deductive manner, to those specific student outcomes mentioned above.

Generating Conceptual Definitions

Upon agreement regarding which generic skill areas will be addressed, the committee proceeded toward the actual generation of conceptual definitions. Each of the representatives on the committee was asked to look at how each of the five skill areas was being promoted in each of the respective academic divisions.

The next step for the committee was the translation of these conceptual definitions into an applicable format.

Operationalizing Conceptual Definitions

"Operationalizing conceptual definitions" means describing, within the parameters of agreed-on definitions, those essential academic abilities which must be demonstrable and measurable by the time of the student's graduation. It was agreed that these student outcomes or performance criteria should be stated in behavioral terms and should specify mediating conditions under which each behavior is to be carried out.[2]

As an intermediate step between the conceptual definitions and the behaviorally stated objectives, the committee chose to illustrate desired student outcomes in general terms. For example, the committee chosen to consider quantification skills as one of the generic skill areas stated general student outcomes in terms of what students must be *able to do* under certain conditions. For example, students will be asked to complete a quantification skills assessment instrument prepared by an appropriate faculty examining committee. They should be able to demonstrate an acceptable level of performance as evidenced by the solution of problems within a specified time frame calling on the abilities enumerated below:

- the ability to measure mass, magnitude, duration of time, number by category on appropriate scale, etc., using the appropriate instruments and units of measure

- the ability to apply computer capabilities to problemsolving (and so on).

From this intermediate step the committee proceeded to enlist the aid of an "appropriate faculty examining committee" to rewrite the aforementioned abilities in behavioral terms. After completion of the behaviorally stated student outcomes, the process of designing and/or adopting an assessment instrument began.

Measuring Student Outcomes

Over the years, many instruments have been designed to measure student achievement. For the most part, these tests measure subject-based knowledge and understanding and do not address such issues as whether students' scores reflect achievement in broad skill areas. More important, such standardized instruments do not necessarily address institution-specific or program-specific student outcomes. Thus, a number of testing instruments were evaluated by the committee. C. Robert Pace, in *Measuring Outcomes of College*, noted several landmark studies worthy of mention here.[3]

The Pennsylvania Study, undertaken in 1928, 1930, and 1932, involved administering to 4,500 graduating seniors from forty-five Pennsylvania colleges a twelve-hour achievement test battery. Sophomores were also tested in 1930 and were again tested as seniors in 1932 in an attempt to produce evidence in support of the obvious notion that the more extensively one has studied in a field, the more one knows about it. As expected, test results substantiated this claim as mean scores in every one of the forty-five colleges improved between sophomore and senior testing. Each college also gained on all or nearly all of the specific subjects measured by the test (except mathematics) regardless of the student's major field of study.

The Tests of General Education produced by the Graduate Record Office were a battery of examinations requiring eight hours to complete. Students were tested in general mathematics, physical sciences, biological sciences, social stuides, literature, arts, effectiveness of expression, and vocabulary. More than 1,000 students from sixteen colleges were tested as sophomores in 1946 and again as seniors in 1948. Mean scores of seniors in all subject areas were higher than those of the same students taking the test as sophomores, thus confirming the results of the earlier Pennsylvania study.

The Graduate Record Examination Area Tests were introduced in 1954 and were designed to test knowledge in the broad categories of social sciences, natural sciences, and humanities with a 3-3/4 hour test battery. These tests represented a significant departure from the Pennsylvania study and the tests of general education. Rather than testing for content knowledge, the tests evaluated students' reading, comprehension, and interpret retation. Again, test scores of 3,035 college seniors in 1954 confirmed the findings of earlier studies regarding students' scoring patterns in their fields of specialization.

The College Level Examination Program was designed to give students college credit for knowledge acquired outside the collegiate context. In order to gather normative data, the tests were given to a national sample of college sophomores and students enrolled in the United States Armed Forces Institute courses. Again, conclusive data were obtained illustrating that the level of student performance was directly related to the number of college courses completed.

In the Undergraduate Assessment Program, Area Tests and Field Tests were closely related to the area tests of the Graduate Record Examination and were initially administered to 47,000 college seniors from 211 colleges during the period 1969-1971. The UAP Area Tests related to students' "areas of interest" and measured achievement in humanities, natural sciences, or social sciences. The UAP Field Tests, on the other hand, measured student achievement in one or more of eleven content-based disciplines.

The Educational Testing Service recently revised the area tests based on data gathered by testing approximately 16,000 seniors in 105 colleges, 2,200 juniors in 57 colleges, 4,200 sophomores in 46 colleges, and 2,800 freshmen in 30 colleges during the period from 1976-1978. Significant differences in student achievement as a function of amount of completed coursework were obtained and proved to be compatible with the findings of all of the previously described studies.

The College Outcome Measures Project, sponsored by the American College Testing Program, is designed to measure communications skills, problem-solving skills, values-clarification skills, functioning within social institutions, using science and technology, and using the arts. Because of the abstract nature of the areas being tested, test materials required students to respond to film excerpts, tape discussions, photographs, music, charts and tables, magazine articles, etc. This battery was particularly timeconsuming and required six hours for students to complete.

The Wayne State Approach

In an attempt to provide a detailed plan of action that might be used by other institutions, there follows the five-stage approach that proved to be effective at Wayne State College. This approach evolved out of the work of the Center Associates, the staff, and the consultants of the Resource Center for Planned Change of the American Association of State Colleges and Universities.

Stage I: Development of Conceptual Definitions

A. *Objective*: To develop and come to initial committee consensus on concise statements defining each of the five skill areas to be promoted, i.e., communicating, quantifying, analyzing, synthesizing, and valuing. These definitions should describe the five skills conceptually as well as operationally and should readily achieve institution-wide understanding.

B. *Rationale*: In order to ensure as wide a basis of consensus and support as possible, it is proposed that the college's faculty senate, futures planning committee, and division heads/deans, be consulted during all stages of the project and that a general education planning committee composed of several students and faculty members from each of the academic divisions be convened. The names of the faculty members should be proposed by each division in consultation with the appropriate constituencies. The names of the students should be proposed by the student senate also in consultation with the appropriate constituencies.

The representatives on this planning committee, as part of their task, will be asked to look at how each of the five skill areas are being promoted in each of the respective academic divisions.

It might also be appropriate that at some future juncture, three additional faculty members be assigned to critique the work of the committee and that, if possible, each committee member and the three additional faculty members be remunerated for contributions beyond duties normally considered part of their workload.

C. *Recommendations*:

1. That each of the respective divisions be allowed to determine the method (either election, appointment, consensus) by which its committee representative is selected.

2. That, as early as possible in the process, faculty members, administrators, students, and standing college committees be apprised of the goals and processes of the project and be continually updated on the project's progress.

3. That the process of working toward the project's goals be recognized as being as valuable as the potential products that might emerge.

4. That administrative intervention be minimized and the perception fostered that the primary administrative function is facilitative in nature.

5. That the general education program be recognized as an institution-wide program to inhibit the development of vested interest groups, turf protecting, and hidden agendas.

6. That all work of the committee be considered tentative and subject to revision by the full faculty.

Stage II: Identification of Observable Performances and Criteria

A. *Objective*: To identify and develop procedures that will elicit student performance indicating the presence of the generic skills and to establish criteria for judging the adequacy of student responses.

B. *Rationale*: This stage requires that levels of performance be established for each of the five skill areas. In order to do so, the committee, in consultation with its constituent faculty and administration, must determine the range of performance levels (or specify desired student outcomes) appropriate to academic standards for undergraduates at the college. There are several issues related to the establishment of levels of performance that must be confronted: (1) Is it feasible (or desirable) to establish minimum levels of performance in these skill areas? (2) Is it feasible (or desirable) to compare input measures to output measures in trying to establish the institution's academic impact upon its students? (3) What implications are extant regarding the possibility of faculty overload in attempting to determine whether and to what extent desired student outcomes have been achieved? (4) How can objectivity be achieved in evaluating student performance levels?

C. *Recommendations*:

1. That the committee recognize and call upon those faculty and staff members having relevant expertise for counsel in determining appropriate performance levels.

2. That the committee recognize the importance of consulting the faculty-at-large concerning the procedures by which student performance evidence be gathered, interpreted, and reported.

3. That, although standardized instruments may be used to assess the various skills, the primary concern of the project is to address institution-specific outcomes within the parameters of the college's general education program.

4. That the most efficient and effective mode of evidence gathering be employed while being mindful of the faculty's workload and day-to-day professional obligations to their classes, with minimal disruption of normal curricular routine, with the least threatening (to both students and faculty members) possible

process of data gathering, analysis, and reporting.

5. That the process employed to determine the extent to which desired student outcomes are achieved be academically defensible and easily understood by the faculty.

Stage III: Assessment of Students and Programs: Design and Implementation

A. *Objective*: To determine procedures for identifing which students at the college will be assessed, to determine the logistics of data collection, to come to committee consensus on the design of the assessment process.

B. *Rationale*: How can we find out whether, where, and to what extent undergraduates at our institution develop communication, quantification, synthesizing, analyzing, and valuing skills? This stage of the project will result in a determination of where, within the curriculum, these skills are being promoted by instructors and manifested by students. Stated in other terms, this stage will result in faculty members, administrators, and students determining what "evidence" will demonstrate that students have acquired, or are in the process of acquiring, each skill. We should look beyond the mere completion of courses and programs for such evidence and should consider such factors as course content, course structure, and course methodology in addition to employing innovative ways of examining existing institutional and program data.

C. *Recommendations*:

1. That each committee member submit a report resulting from examination of each respective division's contribution to the college's general education program in light of the committee-generated definitions of the five generic skills.

2. That the three consulting faculty members submit critiques on all committee work to date.

3. That the committee determine which specific courses in the general education program will be examined and the method by which students enrolled in these courses will be sampled.

4. That a retreat/workshop be scheduled for members of the committee, the three faculty critics, and the faculty senate to discuss the goals and procedures of the project.

5. That institution-wide, general education program statistics be gathered on

a) Enrollment (overall) in selected general education courses during the last three years, including the number of sections offered, total enrollment, student-credit-hour production, and average section size.

b) Enrollment (by semester) in each selected general education course during the last three academic years, including number of sections, total enrollment, average section size, and percentage of each division's general education contribution to the total institutional general education program.

c) Student-credit-hour production in the selected general education courses (by division of instruction) as a percentage of the total general education program semester-by-semester for the last three years.

d) Student-credit-hour production (by semester for the past three years) as a percentage of the total undergraduate production for selected general education courses. This information should include general education student-credit-hour production and general education as a percentage of total undergraduate student-credit-hour production.

e) Undergraduate student-credit-hour production (by division, semester-by-semester) for the last three years.

f) Enrollment (by division, semester-by-semester) for the last three years, including average section size, and general education student-credit-hour production as a percentage of the total institutional general education and as a percentage of the total undergraduate student-credit-hour production.

6. That subsequent to the first semester's final examination period, student performance evidence gathered from selected general education courses be prepared for analysis.

7. That during the following semester, expert committees of the faculty and student body be formed to examine and evaluate student performance evidence.

8. That during the following summer, student performance evidence be analyzed and reports of this evaluation submitted to the project coordinator for synthesis.

9. That as a result of the performance evidence analysis, course syllabi be examined voluntarily by each respective faculty member to minimize the dissonance between stated course objectives and actual course objectives achieved.

10. That additional questions be added to standardized course evaluation instruments (e.g., IDEA, Kansas State University) to generate data on general education course objectives.

State IV: Evaluation–Make Judgments Regarding Student and Program Performance

A. *Objective*: To evaluate the data on student performance and determine the worth of the level of student generic skill attainment in relation to the role and mission of the college.

B. *Rationale*: The assessment design developed by the institution should define the parameters of the evaluation approach. The evaluation component of the project should seek to determine the extent to which the institution has promoted undergraduate competence in each of the five skill areas. In other words, based on the approach to assessment, faculty members should identify and perhaps design measures to demonstrate whether or how well the college has accomplished its goals for student achievement in each of the five skill areas. In this regard they should:

- specify where and to what extent in each academic division, each skill is being promoted
- specify appropriate measures for each skill area, ranging perhaps from unobtrusive measures to formal testing to expert judgment
- design and/or identify the necessary and appropriate administrative structure to carry out the evaluation plan.

C. *Recommendations*:

1. That committee members examine the general education curriculum on a course-by-course basis to determine which courses and constituent students would best be sampled to meet the project's goals.

2. That faculty members teaching general education courses be given the opportunity to volunteer to take part in the evaluation stage of the project.

3. That student performance data on a course-by-course basis be continuously gathered. This evidence may include copies of quizzes, tests, essays, final exams, projects, term papers, and so on.

4. That the anonymity of all students be preserved throughout the duration of the project.

5. That faculty members be assured they will not be judged by the performance of their students.

6. That very close contact be maintained between committee and faculty members teaching the general education courses being examined.

7. That during each semester, faculty members teaching general education courses be encouraged to update course syllabi.

8. That criteria be identified to determine the process by which the "expert" faculty evaluation committees (which will evaluate student evidence) will be formed.

9. That the "expert" faculty evaluation committees identify criteria by which student evidence will be evaluated.

10. That during the subsequent summer, the actual evaluation and analysis of student evidence be carried out, with committee reports submitted to the project coordinator.

State V: Policy Development and Decision Making

A. *Objective*: To remove discrepancies between desired and actual levels of student attainment; to reenter various project stages if definitions need modification or if new assessment/evaluation procedures are to be adopted.

B. *Rationale*: Decisions must be made after the analysis of student performance data regarding the remediation of any discrepancies between the real and the ideal levels of student performance in the various skill areas. Determinations must be made on whether such discrepancies are extant because of a lack of appropriate resources, an inappropriate allocation of available resources, an inappropriate curricular program, or an ineffective curricular design. In order to minimize these discrepancies, institutional policies and procedures must be examined and modified in light of student performance data by institutional consensus.

C. *Recommendations*:

1. That, as a result of the analysis of student performance data, the committee be central in the task of recommending procedures by which the findings of the project are to be presented to the college's various constituencies for discussion.

2. That the college's administration promulgate the perception that the general education program is an institution-wide program and rests in the hands of the entire faculty.

3. That the faculty of the institution be helped to recognize that curricular evaluation (and resultant modification) is a desirable phenomenon in higher education and a necessary element in curricular evolution toward increased effectiveness.

4. That new policies resulting in curricular change are the result of curricular evaluation and *not* the evaluation of specific individuals or divisions of instruction.

5. That a mechanism be employed for the continuing evaluation of the general education curriculum.

Conclusions

The results of the Wayne State efforts to focus on academic competencies in basic intellectual skills, as developed by the APEP project, are presented below. The section includes data on four skill areas: communication, quantification, analysis, and synthesis.

Communication Skills

Definition

Competence in communication, in a broad sense, involves all those procedures allowing an individual to express him/herself effectively. Within this broad range two distinct, clearly related types of competence may be identified. The first is verbal competence, or the ability to speak and write clearly and effectively. The second is nonverbal systems of codification, which range from aesthetic symbols to gesture and digital language systems. It follows that effective communication is a form of human behavior that may be exhibited in a variety of media and contexts.

Desired Student Achievement

Students receiving the baccalaureate degree from Wayne State College should have the following abilities:

- to convey effectively information and ideas in speech, writing, and scientific and technical languages in a clear, coherent, correct, organized, concise, and appropriate manner
- to convey feelings, expressions, and values by means of verbal and nonverbal communication systems
- to receive and understand information, ideas, feelings, attitudes, and values from speech and writing
- to understand ideas, emotions, and values conveyed in art, literature, and music. (Expression through art, literature, and music is a significant form of communication. Although desirable to students achieving proficiency in such methods of communication, it is not something that can be required at the level of general education.)

Performance Level Criteria

Members of the communication skills subcommittee will evaluate the level of each student's performance on a scale ranging from highly effective to not effective. Highly effective performance indicates a superlative level of student achievement while performance judged not to be effective suggests that the student has serious difficulty displaying an acceptable level of achievement.

Data. Number of Course Evaluations = 34; Number of Different Evaluators = 4; Number of Student Evaluations = 868; and Overall Mean of Attained Communications Skills = 3.38.

Mean of Attained Communications Skills. General Education Courses = 3.41, n = 501 and Non-General Education Courses = 3.33, n = 367.

Mean Score of Attained Communications Skills by Course Level. 100 level = 3.38, Student n = 581; 200 level = 3.33, Student n = 254; 300 level = 3.76, Student n = 33; and 400 level = ----, Student n = 0.

Mean Score of Attained Communications Skills by Division. App. Sci. = 3.44, n = 317; HPERA = 3.23, n = 202; Business = ----, n = 0; Humanities = 3.39, n = 235; Ed/Psych = 3.90, n = 11; Math/Sci. = 2.97, n = 48; Fine Arts = ----, n = 0; and Soc. Sci. = 3.74, n = 55.

Mean Score of Attained Communications Skills by Each Evaluator. Evaluator I = 3.18, n = 312; Evaluator J = 3.52, n = 311; Evaluator K = 3.35, n = 234; and Evaluator L = 3.94, n = 11.

Quantification Skills

Definition

Just as college graduates must be able to express themselves in an organized, clear, concise manner using appropriate verbal and nonverbal symbols, they must also be familiar with basic quantitative principles and procedures.

Quantification is the ability to know when it is appropriate to measure, count, and compare and to know how to measure, count, and compare when appropriate.

Desired Student Achievement

Students receiving the baccalaureate degree from Wayne State College should have the following quantification skills:

- the ability to use appropriate instruments to measure number, mass, magnitude, and duration of time
- the ability to determine whether problems can be solved using quantitative techniques
- the ability to record data and to plot and interpret graphs and charts
- the ability to use calculators or computers to solve problems
- the ability to use arithmetic, geometric, algebraic, and statistical techniques to solve problems
- the ability to estimate quantities
- the ability to use basic formulas to convert units of measure
- the ability to determine probability
- the ability to use appropriate measurements in the analysis of art.

Performance Level Criteria

Members of the quantification skills subcommittee will evaluate the level of each student's performance on a scale ranging from highly effective (5) to not effective (1). Highly effective performance indicates a superlative level of student achievement while performance judged not to be effective suggests that the student has serious difficulty displaying an acceptable level of achievement.

Data. Number of Course Evaluations = 33 (some with combined multiple sections); Number of Different Evaluators = 3; Number of Student Evaluations = 1526; and Overall Mean of Attained Quantification Skills = 3.57.

Mean Score of Attained Quantification Skills. General Education Courses = 3.54, n = 1302 and Non-General Education Courses = 3.74, n = 224.

Mean Score of Attained Quantification Skills by Course Level. 100 level = 3.59,Student n = 1508; 200 level = 0, Student n = 0; 300 level = 2.00, Student n = 18; and 400 level = 0, Student n = 0.

Mean Score of Attained Quantification Skills by Division. App. Sci. = 3.35, n = 48; HPERA = ----, n = 0; Business = 3.62, n = 1334; Humanities = ----, n = 0; Ed/Psych = ----, n = 0; Math/Sci. = 3.22, n = 144; Fine Arts = ----, n = 0; and Soc. Sci. = ----, n = 0.

Mean Score of Attained Quantification Skills by Each Evaluator. Evaluator A = 3.56, n = 497; Evaluator B = 3.57, n = 507; and Evaluator C = 3.58, n = 522.

Analysis Skills

Definition

Analytical skill involves identifying the parts of a whole and the relationships of those parts to the whole. It includes the ability to identify the essential components of ideas, events, problems, and processes; to draw logical deductions about those components; and to recognize the limitations of these deductions. Because most problems are not isolated from one another, analytical skill includes the ability to establish relationships, when they exist, between problems.

Analytical competence also involves the ability to think critically and to solve problems.

Desired Student Achievement

Students receiving the baccalaureate degree from Wayne State College should have the following analytic skills:

- the ability to recognize what facts or assumptions are essential to or implicit in a thesis, claim, explanation, argument, or belief
- the ability to recognize and distinguish between the structural elements of oral or written communication, including introduction, premise, conclusion, and illustration
- the ability to devide compound or complex claims into simpler components
- the ability to detect relationships between ideas presented
- the ability to perceive connections between facts, events, or elements presented
- the ability to recognize how a theory, process, work of art, or scientific concept explains or orders diverse elements
- the ability to reason inductively and deductively
- the ability to assess the validity of inductive and deductive arguments
- the ability to detect cause and effect
- the ability to distinguish between the important and unimportant details in a passage or presentation
- the ability to detect faulty reasoning
- the ability to detect inconsistency and incoherence
- the ability to gather sufficient information to provide a data base for analysis
- the ability to distinguish between facts and assumptions.

Performance Level Criteria

Members of the analysis skills subcommittee will evaluate the level of each student's performance on a scale ranging from highly effective to not effective. Highly effective performance indicates a superlative level of student achievement while performance judged to be not effective suggests that the student has serious difficulty displaying an acceptable level of achievement.

Data. Number of Course Evaluations = 44; Number of Different Evaluators = 2; Number of Student Evaluations = 756; and Overall Mean of Attained Analysis Skills = 2.88.

Mean Score of Attained Analysis Skills. General Education Courses = 2.82, n = 476; and Non-General Education Courses = 2.99, n = 280.

Mean Score of Attained Analysis Skills by Course Level. 100 level = 2.63, Student n = 498; 200 level = 3.09, Student n = 93; 300 level = 4.53, Student n = 19; 400 level = 2.73, Student n = 136; and 500 level = 3.80, Student n = 10.

Mean Score of Attained Analysis Skills by Division. App. Sci. = 1.59, n = 37; HPERA = 2.40, n = 115; Business = 3.00, n = 3; Humanities = 3.04, n = 237; Ed/Psych = 3.26, n = 90; Math/Sci. = 2.71, n = 56; Fine Arts = ----, n = 0; and Soc. Sci. = 2.89, n = 218.

Mean Score of Attained Analysis Skills by Each Evaluator. Evaluator D = 2.95, n = 409 and Evaluator E = 2.80, n = 347.

Synthesis Skills

Definition

Synthesis means putting together separate elements into a usable or satisfying whole such as a theory, process, work of art, or physical construction. It is the organization of phenomena, events, ideas, or beliefs to yield a new order not previously apparent. (A theorist, for example, seeks to pull together seemingly disparate data into a comprehensible whole.)

Synthesizing skills are those which help an individual integrate basic knowledge, feelings, or the like. At all levels of every field of study and work this skill allows individuals to create innovative combinations of ideas, objects, processes, and people. Students with highly developed synthesizing skills are adept at putting apparently unrelated facts into a common frame of reference. They are able to find unifying themes or common threads that enable them to deal effectively with diffuse bodies of information. They not only possess a facility for constructing a general framework to accommodate whatever information they have, but they can move beyond that information, applying its principles to new situations or other bodies of information.

Desired Student Achievement

Students receiving a baccalaureate degree from Wayne State College should have the following synthesizing skills:

- the ability to organize related facts, ideas, and beliefs into a coherent framework
- the ability to introduce new facts, ideas, and/or beliefs into an existing framework
- the ability to extrapolate from known data to unknown
- the ability to recognize connections between theory and practical application
- the ability to use data and methods of different disciplines to understand a problem.

Performance Level Criteria

Members of the synthesis skills subcommittee will evaluate the level of each student's performance on a scale ranging from highly effective to not effective. Highly effective performance indicates a superlative level of student achievement while performance judged not to be effective suggests that the student has serious difficulty displaying an acceptable level of achievement.

Data. Number of Course Evaluations = 58; Number of Different Evaluators = 3; Number of Student Evaluations = 1247; and Overall Mean of Attained Synthesis Skills = 3.62.

Mean Score of Attained Synthesis Skills. General Education Courses = 3.79, n = 584; and Non-General Education Courses = 3.46, n = 663.

Mean Score of Attained Synthesis Skills by Course Level. 100 level = 3.59, Student n = 780; 200 level = 3.49, Student n = 225; 300 level = 4.01 Student n = 162; 400 level = 3.42, Student n = 74; and 500 level = 3.67, Student n = 6.

Mean Score of Attained Synthesis Skills by Division. App. Sci. = 4.18, n = 130; HPERA = 3.58, n = 109; Business = 3.11, n = 207; Humanities = 3.95, n = 216; Ed/Psych = 3.35, n = 188; Math/Sci. = 3.22, n = 95; Fine Arts = ----, n = 0; and Soc. Sci. = 3.79, n = 302.

Mean Score of Attained Synthesis Skills by Each Evaluator. Evaluator F = 3.58, n = 848; Evaluator G = 3.58, n = 335; and Evaluator H = 4.50, n = 64.

Notes

[1] Paul L. Dressel, "Methods of Evaluation and Research," *General Education* ED. Louis B. Mayhew (New York: Harper and Bros. Publishing, 1960), pp. 167-168.

[2] Paul D. Plowman, *Behavioral Objectives* (Chicago: Science Research Associates, 1971), p. xxvii.

[3] C. Robert Pace, *Measuring Outcomes of College* (San Francisco: Jossey Bass, 1979), pp. 10-47.

Written By

Bart Brady-Ciampa

Institutional Contact

Donald W. Whisenhunt, vice president

Western Carolina University (NC)

Institutional Profile

Western Carolina University (WCU), a state-supported coeducational institution established in 1889 and located in the scenic Appalachian mountain ranges at Cullowhee, North Carolina. The university consists of the main campus in Cullowhee and resident credit centers in Asheville and Cherokee. It includes six schools offering degrees at the bachelor's, master's, and education specialist levels.

Background

The meaning of general education was the question at the heart of the Academic Program Evaluation Project (APEP) from the outset. Western Carolina University had moved very rapidly in the late 1960s and early 1970s from the status of a small college known primarily as a teacher training institution to the position of a regional university with five undergraduate schools and a graduate school. As the curriculum developed to accommodate new or expanded fields of study, the general education requirements had become a subject of heated debate. At the time the APEP idea was proposed, a WCU faculty committee had just completed a two-year study of the general education program and had proposed an extensive revision. Having taken part in the AASCU forum in which the concept of the project evolved, the vice chancellor foresaw that APEP could significantly affects for the reconsideration of general education.

The APEP design called for defining and measuring five intellectual skills; the WCU general education plan suggested ten content areas and seven skills as appropriate to a general education curriculum. WCU's miniproposal to AASCU to participate in APEP emphasized the value of such an evaluation process in clarifying issues and directions involved in redefining general education. It was thought that APEP participation would reinforce interest in developing course descriptions and requirements that included considerations of accountability and measurement of outcomes. The opportunity to engage in pre- and post-outcome measures while the redesign of general education was in progress appeared opportune for estimating the relative effectiveness of old and new approaches to course and program objectives. The project was seen as a way to develop a sound basis for describing objectives in behavioral terms as well as for learning how to develop measures that could assist the faculty in judging whether intended effects had been achieved. The project design also appeared to afford ways to identify more clearly what program elements might contribute to desired general education outcomes. Thus, despite the integrity of APEP and the entirely separate function of its steering committee, nearly every early decision relating to the project was colored by its inseparable relationship with the general education reassessment.

The fifteen-member APEP steering committee was selected by the vice chancellor for academic affairs, who served as project director, and the director of special academic programs, who was project coordinator. The active involvement of the chief academic officer of the university was seen as essential to giving the project high visibility and to assuring faculty participants of the project's high priority within the administration. Additionally, as an expert in the field of measurement and evaluation, the vice chancellor was uniquely suited by training and experience to provide strong leadership for the project. Another vital consideration was to name a coordinator with a record of success in conducting long-term, multifaceted projects. Finally, the whole group had to be task oriented and capable of sustained commitment.

Selection was designed to provide representation from all five undergraduate schools. Individuals were selected on the basis of their record of leadership and credibility among their peers. Although all of the committee, including two replacements, were among the busiest and most committed members of the faculty, all but one of the rest of the committee stayed with the project for two years. The decision to recognize and cope with the problems of overload inherent in selecting the busiest faculty members has proven positive, primarily because these individuals have infused the effects of the APEP process into the continuing, wide debate on implementation of a new general education program. The indirect impact was significant throughout the two years of the project.

For the first year of the project, both the director and coordinator sought ways to expand the membership of the steering committee, with the intention of achieving more widespread involvement

74

and, hopefully, better acceptance of the projected program evaluation. However, the committee feared that "going public" with APEP would exacerbate and confuse the already heated general education debate.

The proposed general education plan stated that "general education should . . . foster intellectual self-reliance and ability through extensive practice in reasoning and critical thinking; in experiencing the varied methods by which truth and knowledge are acquired and applied in organizing, analyzing, and synthesizing data, knowledge, and ideas; and in communicating the results thereof." The general education committee had said in its report that "no course is to be accepted for the program unless its plan includes adequate and evaluative exercises by each student of written communication in the English language, oral communication, the use of scientific method, the use of critical and analytical thinking, the use of logical reasoning, and the making of value judgments." Clearly, APEP and general education could reinforce each other. The question was how to maintain the distinctly separate character of APEP long enough to produce meaningful definitions and measures of intellectual skills while the controversy over the content and structure of general education remained the central issue on campus. Repeatedly, the committee reached the same conclusion: keep the APEP committee small, avoid publicity, and keep working. Ultimately, only a few individuals were added to the committee after initial appointments. Given the continued controversy over general education on campus and the considerable success of APEP in achieving its goals without arousing faculty opposition, the decision not to expand the committee and to maintain a low profile appears to have been a good one for our particular circumstances.

Implementation

Project Design

The project called for assessment of four skills, with a fifth skill--valuing--optional. The WCU committee agreed to delay consideration of the skill of valuing and to begin the task of defining analysis, synthesis, communication, and quantification. Initial discussions led to comments about the futility of reinventing the wheel and suggestions that the committee adopt or adapt existing definitions or hierarchies such as Bloom's taxonomy. But subsequent reviews of the literature proved more helpful in providing a background for discussion than a basis for agreement.

Management of the task required maintaining a careful balance between providing adequate background information and avoiding an information overload. Faculty members were asked to review key materials in the field, but for the most part the common understandings that had to be developed came from hearing presentations brought to the group by the project director, outside consultants, and faculty experts. Ultimately the committee split into subgroups, developed institution-specific definitions of the skills, and presented them to the full committee for debate, revision, and approval, a process that occupied nearly five months. The time devoted to this stage of the project was well spent. During these months the committee gradually overcame initial reluctance to wrestle with the hard questions of what they actually meant by analysis, synthesis, communication, and quantification.

The major difficulty in defining analysis was in determining the appropriate sequential order of subskills and in clearly delineating at what point the skill of analysis merges with synthesis. Discussions of synthesis led to use of terms such as *chaos*, which in turn brought about lengthy consideration of the nature of order itself and the extent to which the committee believed synthesis is a process of perceiving as well as creating order. The definition of *communication* proved more difficult to contain than the others, as the term seemed to include receptive as well as expressive skills and, under examination, expanded rapidly with the incorporation of nonverbal languages, including such disparate forms as the arts and computer technology. Finally, in defining quantification the committee found that in addition to being a process for identifying and manipulating magnitudes and symbolic relationships, it, too, was a means of communication employing mathematical and graphic symbols.

Agreement came slowly, but with it came two very positive side effects. The committee expressed a sense of exhilaration at having spent time thinking and talking with colleagues about what education really meant. And the group achieved a breakthrough of understanding when, after months of frustrating, time-consuming effort that seemed doomed to failure, they realized that the process of defining skills and designing measures was far more important and useful than any product that might result from the project. That same realization rescued the work from low moments throughout the two and one-half years the committee worked together. At the end, the process had been significant both in individual professional development and in the multiple effects on campus of these faculty leaders.

As the committee began trying to develop per-

74a

formance measures, early definitions that had seemed acceptable proved unworkable. The first of many loop-back cycles occurred as the committee rewrote skill definitions to include specific, measurable subskills.

Often the main obstacle to progress in the discussions was simply the lack of a commonly understood term. When the committee seemed destined for perpetual disagreement in its effort to work from definitions of generic skills to observable indicators of these skills, the missing term emerged as *enabling skills.* The subskills came quickly after that point. A similar semantic breakthrough resulted from the use of *indicator* in place of *performance measure* or *test item.* The committee also found it more reasonable to develop an "assessment design" than an "evaluation plan." The search for a common vocabulary that was not only clear but also comfortable was a continuing challenge to managers of the project.

In an effort to give the committee some kind of common background in evaluation technique, two faculty members with specialties in the field were asked to present a workshop for the group. The workshop proved helpful in supplying a common vocabulary. Even more helpful, the workshop presenters were added to the committee. At the end of the first semester, small task groups agreed to work during the summer selecting or developing appropriate measures of the intellectual skills the committee had defined.

The plan for the summer called for a review of commercial test instruments by two of the committee's experts on measurement with recommendation for use or adaptation. The materials were also to be employed as models for the committee to imitate in writing homemade indicators. Having agreed that the scope of the design would extend only to general education, the summer task group began creating a pool of test items and behavioral indicators, all of which were outlined according to subskill, content/context, cue to elicit the subskill, tools and conditions, and performance criteria. By the end of the summer, the group could present a selection of more than seventy-five possibilities to the steering committee.

The major issue to emerge from the summer's work was not a matter of what measures to employ or what sample to assess, but rather an issue that had been too quickly dismissed months earlier: the purpose of the project. It had first appeared that by taking part in the national project WCU was simply adopting the purpose of the national project as its own. An early statement of purpose required no more than twenty minutes of discussion. The committee would "assess the skill levels of students in four intellectual skills identified as common objectives of baccalaureate education, identify and describe program elements in general education that are probable contributors to acquisition of these skills, and on the basis of that data evaluate and possibly revise the general education program." It seemed simple; nonetheless, it proved to be far beyond the scope of the design that emerged at the end of the first year's work.

The question of purpose provided the major area of debate as the full committee resumed meeting in the fall. What could the project realistically hope to measure and to what extent could data be regarded as sufficiently reliable for making program and policy decisions? The committee began to believe that such causal connections between program elements, student intellectual behaviors, and test outcomes as might exist could not be clearly demonstrated.

The APEP paradigm that provided an outline for the project evolved and changed as ten institutions tested its effectiveness. The WCU project found that the paradigm had to be seen as iterative rather than linear or totally sequential. Loop-back cycles were frequent, beginning with Stage II, in which the task of devising applications for abstractions had given the project a firm base in reality. Definitions, statements of purpose, and long-range plans for policy and program change gave way to the major question of what could be observed or measured as an indicator of the generic skill and then confidently related to a course or program element. At that point the committee reached a major agreement, that the project design was to be formative, descriptive, and hypothesis-generating rather than summative, causative, and hypothesis-testing. The new purpose statement reflected that change:

> The purpose of the Academic Program Evaluation Project at Western Carolina University is to assess the skill levels of students in the four intellectual skills identified as common objectives of the general education portion of a baccalaureate education and to identify, describe, and increase awareness of program elements in general education that are probable contributors to acquisition of these skills. We intend to carry out such evaluations as are possible within the time and resource constraints of the project and, dependent upon the adequacy of the data, to make such policy decisions as seem justified. We expect, however, that it will be necessary to use an iterative approach and to employ more complex evaluative designs later to get at some of the intricate causal connections that WCU will want to consider.

Test Design

Early discussion with the national director of the project about choosing the final evaluation design resulted in a decision to bring in a consultant from the Educational Testing Service to examine the committee's work and to suggest future directions. Having had the advantage of an on-campus expert as director of the project, the committee had had little need to bring in outside assistance earlier. At this stage, however, it was useful to have outside assurance that the design and indicators were basically valid. The committee then moved toward a design and the choice of a student test sample. The committee believed that any survey of course content or any measurement of skill development in most general education classes would lead to immediate faculty resistance and suspicion. Consequently, the primary test sample was chosen from the general psychology course, for which a research component is required. Any student could meet that requirement by serving as a research subject for up to three hours of testing. The APEP battery was condensed to a three-hour core and a separate set of open-ended behavioral measures. One faculty member conducted the test in the spring and another the following fall. Two graduate students were also employed to assist with the project.

The evaluation plan involved the assessment of first-semester freshmen and second-semester sophomores. A total of seventy-eight sophomores were tested, fifty-six of whom had one all of their college work at WCU and twenty-two who were transfer students. In addition, sixty-two first-semester freshmen were assessed. Both groups were representative of the entering student population at WCU based on a comparison of SAT scores. Two other sets of students participated. One set comprised three groups of upperclassmen ($n = 14$ in 3 groups) and three groups of first semester freshmen ($n = 13$ in 3 groups). These students took a multiple-choice test of synthesis skills and engaged in a videotaped, group problem-solving task. The second set was made up of twenty upperclassmen and sixteen first-semester freshmen. These individuals delivered an impromptu speech and took the Huey-Johnson Listening Test.

Testing Measures

There were nine tests and exercises administered to the freshmen and sophomore groups during the three-hour testing session. Some of the tests had been developed by ETS for use in APEP-type projects. These were Communication--Form 14; Mathematical Thinking--Form 27; Analysis--Form 34; and Synthesis--Form 44. Other tests used in this testing session were assembled by members of the APEP Steering Committee. These were a 23-item mathematics test, a 33-item analysis test, a 30-minute essay, and two interpretive analysis exercises. A standardized test--the Nelson-Denny Reading Test of comprehension, vocabulary, and reading rate--was also given. A brief description of each testing instrument appears below.

Analysis

Test: ETS: Analysis

Description: An ETS-developed, multiple-choice pilot test of the ability to judge the logic of an argument in textual material (nine items).

Test: WCU Analysis

Description: A locally developed objective test that included items regarding the relative difficulty of proof, analysis of assumptions, and the distinction between fact and opinion (thirty-three items).

Test: Analysis Essay

Description: A series of eight short-answer questions about textual material.

Synthesis

Test: ETS: Synthesis

Description: An ETS-developed pilot test of the ability to make judgments about information provided in tables and graphs (twelve items).

Quantification

Test: ETS: Mathematical Thinking

Description: An ETS-developed pilot test of basic mathematical knowledge including fractions, percentage, simple algebra (fourteen items).

Test: WCU Test of Quantification Skills

Description: A multiple-choice test of basic mathematical knowledge such as addition of fractions, calculations, simple algebra and geometry, and basic set theory operations (twenty-three items).

Communication

Test: ETS: Communication

Description: An ETS-developed, multiple-choice test requiring students to judge the relative clarity and completeness of written passages (twelve items).

Test: WCU Essay

Description: An open-ended writing task in which students prepared a paragraph on one of three topics.

Test: Nelson-Denny Reading Test

Description: A standardized test that yielded scores for comprehension, vocabulary, and reading rate (thirty-six comprehension items, 100 vocabulary items).

Other data-collection exercises involved additional freshmen and upperclass groups. One set of students took the Synthesis test--Form 44 and then participated in a small group problem-solving task. Each small group devised a strategy for determining student representatives to a standing government committee on student problems. The strategy of each group was evaluated using a rating scale that assessed task identification, organization, logical agenda, participation, time management, and achievement of purpose. The remaining students who participated in the project delivered a five-minute impromptu speech and took the Huey-Johnson Listening Test. The purpose of these tasks, the problem solving and the speech, was to evaluate synthesis and communication skills using tasks which were more likely to arise in the students' postcollege careers.

The evaluation procedure involved administering the three-hour battery of tests to a group of seventy-eight sophomores, the group problem solving task to three small groups of upperclassmen (total *n* = 14), and the impromptu speech and listening test to another group of sophomores (*n* = 20) during the spring 1981 semester. The same tests and exercises were administered to groups of freshmen at the beginning of the fall semester 1981. The corresponding group sizes for the tasks named above were sixty-two for the core battery, thirteen for the group problem solving task, and sixteen for the impromptu speech.

With testing underway, the faculty agreed to begin setting performance standards by taking the test battery themselves. Setting performance standards and evaluating test results raised an immediate fundamental question: What were we measuring with our tests? The committee's original assumption was that the battery of tests would measure four generic skills rather than ones that simply accrue through general maturation or postadolescent interaction with the environment. To test that assumption, the committee first asked whether the test measured skills possessed by any college-educated person. A quick look at the performance of the ten faculty members on the test battery suggests they did not. Faculty performance did not exceed 90 percent correct on any test. Median percentage correct on the measures was 80 percent. Thus, even a select group of faculty members was unable to show "mastery" level of performance on this particular set of measures of thinking and communication skills.

The committee then asked whether the tests measured four distinct skills or sets of skills. This question was addressed with a factor analysis of the performance of the non-transfer students on the test battery. (Transfer students were omitted because information on certain variables was not available.) Factor analysis, a statistical procedure for sorting out patterns of individual differences in test performance, determines the degree to which the skills underlying responses to one or more tests correspond to the skills underlying responses to one or more other tests. Hypothetically, if four generic skills were being measured by the test battery, four different "factors" should have been obtained in the factor analysis. For example, a "quantitative" factor defined by scores on the two quantitative measures should have differed from an "analysis" factor defned by the two tests of analysis. Student SAT scores and high school rank were also included in this analysis in order to determine whether the battery measured anything different from general intellectual ability and academic aptitude.

Outcomes

The results of the factor analysis demonstrated that, to a large degree, the APEP battery measured two skills rather than four. These two factors accounted for 48 percent of the common variance in the scores. The remaining variance was distributed across several factors, most of which accounted for aproximately 5 percent additional variance. The first factor was defined by "loadings" (roughly, correlations between the tests and the factor) of all the tests except the WCU essay. This factor closely resembles what is usually called "general intelligence." The presence of high loadings of SAT scores on the factor supports this contention. The second factor was defined by those tests having a substantial verbal component. The only measure of communication not loading on this factor was the ETS Communication Scale.

One caveat is in order. The internal consistency reliability estimates on the tests tended to be rather low. Thus, a considerable portion of the variation in performance is a result of measurement error. This, combined with the dominance of the general intellectual factor in the factor analysis, indicates that only a few of the differences in student performance can be attributed to uniquely defined generic skills. Except for the WCU analysis test and the Nelson-Denny Reading Tests, performance variability was largely explained through common factor variance and error variance. The WCU Analysis test, in particular, may be worth a further look for evaluation purposes. It was designed to assess a student's ability to judge the difficulty of proving an assertion, to discern fact from opinion, and to determine the importance of data to an argument. These are all skills that most college instructors would consider important and "teachable." They may represent a level of skill intermediate to general intelligence and basic facts.

The small amount of unique variance in the tests prompted the APEP committee to question the validity of attempting to make any evaluation of intellectual skill development in to the collegiate experience. However, after further debate, the group agreed that college should influence the intellectual development of students; therefore, evaluating performance on some specified measure of general intellectual development would be desirable in an evaluation of college outcomes. Referring to *Evaluation Research in Education* by Cooley and Lohnes, which promotes the idea of looking at general intellectual skills in evaluations of educational programs, the APEP committee turned its attention to measuring general intellectual development by generating scores (factor scores) on each of the two factors produced through the analysis. The two new variables, general intellectual ability and verbal skills, were also used in the evaluation paradigm.

The committee next asked what student characteristics were related to performance on the tests. The answer was that the most important student characteristic pertinent to the evaluation intentions of the committee was class level. Did students who had taken general education courses (sophomores) do better on the tests than students who had not (freshmen)? The sophomore group was broken down into students who had spent their first two years at WCU and those who had transferred to WCU from other institutions. These results showed that nontransfer sophomores consistently outscored freshmen (except on the written essay) and that transfer sophomores generally scored at an intermediate level. One means of assessing the size of these differences is in terms of "effect size"--the difference in average (mean) performance as a ratio of variability in scores (the standard deviation of scores). The "effect size" measure ranged from a low of 0.22 to 0.55 for nontransfer sophomores compared with freshmen. Similar effect sizes were noted on the group problem-solving task (0.29), on the Huey-Johnson Listening Test (0.09), and on the impromptu speech (0.31). It is important, from the evaluator's perspective, that these effects occurred despite a lack of difference on SAT scores and a difference in favor of freshmen of high school rank. The consistency in direction and magnitude of the effect sizes associated with the more general group problem solving and impromptu speech tasks seems to indicate that differences on these tasks have the same origin as for the performance differences on the core battery.

Another way of looking at the relationship between student characteristics and student performance is with the use of a statistical technique called "regression analysis." Regression analysis in this application allows an assessment of the degree to which a set of student characteristics "account" for variability in student performance when other characteristics are statistically controlled. The central question in this analysis was how much of the difference among students on each test could be accounted for by class level (freshman/sophomore) when entering ability (SAT scores and high school rank) is controlled. (A regression approach was not used with the group problem solving and impromptu speech data because of the small group sizes.) This procedure was conducted by taking each of the measures in the core test battery and the two factor scores and using each as a dependent variable. The SAT variables and the high school rank variable were entered first into a prediction equation. Any additional variation "accounted" for with the inclusion of class level was interpreted as reflecting the impact of the first two years of college. The changes were rather small in magnitude but consistent with other reported research on the impact of college. From the results from the regression analyses, it appears that the first two years of college at WCU produce differences in intellectual skills whose magnitude confirms the findings from similar evaluations in other educational programs. One of the R-squares of particular interest is the WCU Analysis Test. It was noted in the factor analysis results that this test contained more reliable unique variance than any of the other tests in the core battery. The R-square change associated with this test was also larger than any other. These findings suggest that the APEP approach of focusing on specific, generic skills has merit but that better tools are needed for assessing these attributes.

The evaluation of student performance can be summarized as follows:

- Rather than measuring four distinct generic skills, the tests employed in the APEP battery apparently assessed general intellectual ability and verbal skills for the most part. There was a unique contribution from each test to student performance beyond that attributable to measurement error, but this tended to be small. The tests with the largest portion of unique reliable variance were the WCU Analysis Test and the Nelson-Denny Reading Comprehension Test. In the analysis of freshmen/sophomore differences, performance on the analysis test appeared to be influenced more by the "college years" than was the reading comprehension test.
- Sophomores consistently outscored freshmen on the APEP tests (except for the communication essay). The

exact reason for this difference cannot be determined within the scope of the APEP evaluation design. However, the direction and magnitude of the observed differences are consistent with published reports of other program evaluations. Something associated with the "college years" produced increased levels of performance on practically all measures of generic skills and the composite measures of intellectual ability and verbal skill.

Perhaps the same growing sense of ease and commonality of purpose among the committee members that had encouraged them to take their own tests made development of a program-portrayal document a simpler task than had been anticipated. Careful wording, rewording, and effective group interaction produced a final document that most believed would elicit the kind of information the committee wanted. The question was how and when to use the questionnaire to avoid confusion with an assessment of general education course content being conducted by the faculty senate that same semester. Strategy was simple but essential. The committee agreed to delay the project until fall; distribute copies to the heads of all departments in which general education courses are offered; give them a brief summary of APEP's purpose and activities; ask these department heads to distribute the questionnaires at their first departmental meetings; and have faculty members return the questionnaires, unsigned, directly to the APEP committee. The amazingly high 70-percent return and the absence of any questions or apparent resistance seemed to demonstrate that both the document and the method were sound. It may also support the contention that being attuned to the temper of the campus is a key to success in evaluation, and a committee of faculty leaders is most likely to be best attuned.

APEP committee members were asked to set standards for the program portrayal performance by completing a questionnaire for each of the major subcategories of general education courses. The committee members were instructed to indicate what they considered an ideal level of emphasis on each generic skill to be in each subcategory. The average committee response was used as a standard against which faculty responses could be judged. In nearly every case, the general education faculty's reported emphasis on the generic skills in courses is close to or considerably above the ideal set by the APEP committee. Thus, the general education faculty members appear to perceive themselves as currently placing a great deal of emphasis in the conduct of their courses on the generic skills of analysis, synthesis, communication, and quantification. An on-going aspect of the program description is the use of a slightly revised program portrayal instrument designed to assess student perceptions of faculty emphasis on generic skills and student expectations for workloads.

Conclusion

From the outset, APEP was viewed as a project that would address multiple policy issues on the WCU campus, influencing policy making at all levels in a variety of ways, both formal and informal. As the project neared completion, it was clear that APEP had indeed had a significant impact on policy making. One of the most interesting effects has been that many of the APEP steering committee members have moved into key positions in the university's governance and administrative structure. For example, one of the APEP members is now serving as chairman of the faculty: another chairs the influential academic affairs committee; another has assumed the leadership of the English departments still another has moved into the position of director of international instructional programs; and three APEP committee members serve on the faculty senate. It is equally interesting to observe the degree to which APEP experience has been reflected in positions these individuals have taken on issues before the university. Clearly, the experience of examining and attempting to measure and evaluate student intellectual performance has been and will continue to be significant in policy matters that come before these several individuals and the groups they serve.

The APEP group has also influenced the formation, composition, and efforts of a special monitoring committee established as a result of the acceptance of a new plan for general education. As part of a national meeting of all APEP schools, held on the Western Carolina campus, the WCU steering committee offered to stage an actual case study for discussion by the member schools. The question posed consisted of two parts, both suggested by the actual evolution of the project. First, the skill of valuing had been left for later discussion: second, the merger of APEP's work with the on-going reorganization of general education had yet to be accomplished. Could APEP provide the process by which the campus community might grapple with defining and measuring just one of the seven skills proposed for inclusion in all general education courses--the skill of valuing? From that discussion and subsequent steering committee meetings came the decision that the APEP committee would complete the evaluation project using the four skills originally chosen, present the results to appropriate campus governance bodies, and would then propose that any academic program evaluation project that might be undertaken as part of the new general education monitoring system be designed with advice and assistance from the APEP committee. The committee suggested that the first and most appropriate vehicle for providing that assistance would be a joint meeting with the "General Education Monitoring" (GEM) Committee.

The work of the APEP committee was also instrumental in gaining support for expanding the work of the Learning Resources Center (LRC), a unit established under another grant. The linkage between APEP and the work of the LRC has given support to the task of evaluating programmatic and student outcomes as well as the effort to link evaluation components into locally derived innovations. Certainly the success of APEP members in developing and field testing instruments of their own creation has supported the policy position that LRC should be provided with resources for encouraging faculty development of APEP-like measurement instruments.

Another interesting outcome of APEP is its reinforcement of an existing policy that supported provision of funds for this university to engage in exchange projects with other institutions. APEP, although only one of several current intercampus exchange activities that have borne considerable fruit, was different in that it involved so many other institutions over such an extended period of time. The success of this extended project has given considerable support to the argument for continuation and expansion of funding the interuniversity exchange efforts.

Several policy issues currently before the university will continue to feel, at least indirectly, the influences of APEP experience. For example, implementation of the content and distribution requirements of the new general education program is now underway. The issue of whether the distribution requirements should contain courses emphasizing analytical and synthesizing skills was settled affirmatively, and APEP must receive some of the credit. Similarly, the faculty has shown considerable interest in identifying resources for teaching such generic skills as analysis and synthesis. Continuing discussion about which methods might be best, which might be used in which disciplines, and whether indeed some of these skills can be taught in courses as currently offered argues for special attention to these issues and development of basic policies on resource allocation for examination of these and related issues.

Finally, and perhaps most important, APEP has stimulated renewed interest in program evaluation, has increased awareness of instrumentation issues in instructional evaluation efforts, and has increased sensitivity to the problems involved. At the same time the question of resource allocation and the issue of directions to be taken in internal evaluation efforts have been placed high on the agenda of both the administrative and faculty governance structures. At least one major policy decision has already been made as a direct consequence of the project. The office of the vice chancellor for academic affairs has adopted a long-range policy on evaluation calling for support of evaluating measurement instruments, whether locally developed or nationally available, and for implementation of a longitudinal study of students as they move through the new general education program. Comparisons between the new and old general education programs made possible by APEP's initial work will provide a valuable data base. Additionally, the initial evaluation design can be continuously reevaluated, leading eventually to sound and accepted bases for examining other academic issues sure to arise in the future. The end of APEP is only the beginning of academic program evaluation at Western Carolina University.

Written By

Marilyn Jody, Bruce Henderson, and Robert Pittman

Institutional Contact

Yvonne Phillips, interim vice chancellor for academic affairs

Western Kentucky University

Institutional Profile

Western Kentucky University (WKU), a publicly supported institution established in 1906 as a two-year state normal school, presently offers a comprehensive range of undergraduate and graduate progrms within its four undergraduate colleges and graduate college. Total enrollment is nearly 13,000, with a full-time equivalent student enrollment of about 10,000. Of the students, 85 percent are undergraduates, nearly 85 percent are Kentucky residents, and more than 50 percent are women. The more than 650 faculty members represent thirty-six disciplines and specialties. The range of academic offerings includes technical programs at the associate and baccalaureate degree level; traditional baccalaureate programs in the arts and sciences, education, agriculture, business, health, and other broad fields; special professional and preprofessional curricula; and a number of graduate programs at the master's and specialist degree levels.

Background

Inclusion as one of ten AASCU institutions in the Academic Program Evaluation Project (APEP) provided WKU with a unique opportunity to examine even more closely the results of its instructional programs. Already a national leader in the evaluation of teacher education programs, WKU had more recently implemented a system for periodically reviewing all major programs. Neither of these efforts, however, assessed student attainment of the basic intellectual skills. For some time concern had been expressed among the faculty that students who enter without adequately developed skills in the area of written communication, quantification, and critical thinking are not provided ample opportunity within the baccalaureate program for developing each of those skills to the desired level and sustaining or improving them through continuous use. Thus, APEP was viewed as a means for beginning to examine that development.

APEP at Western Kentucky University was undertaken to evaluate student performance in four areas of intellectual competency--communication, analysis, synthesis, and quantification. A simulated pre- and post-test design was planned: a stratified (by college) random sample of graduating seniors and a group of incoming freshmen (matched to the seniors in terms of scores on the ACT mathematics and English scale) were to be tested in each of the competency areas and their scores compared. By using this procedure, the contribution of the program of study to the development of the intellectual skills was to be estimated.

The initial APEP organization consisted of the vice president for academic affairs, who served as project director; the associate vice president for academic affairs, project coordinator; and the general education committee for the academic council, which also functioned as the APEP committee during the first year of the project. This standing committee consists of fourteen members elected annually from the membership of the academic council: two faculty members from each of the four undergraduate colleges, two representatives from the academic services area, and two students. In addition, two administrators serve as ex-officio members of the committee.

Skills Definitions

The project director and project coordinator (an ex-officio member of the committee) held an initial meeting with the APEP committee to provide information, answer questions, identify the tasks to be accomplished, and discuss potential benefits of the project. The need for thorough understanding of the project and some initial reluctance to accept the challenge was exhibited during the first several of the weekly meetings. To assist in that regard, the project coordinator drafted a WKU-APEP "Focus and Goals" statement (end of chapter), which was approved by the committee and distributed for information and comment to the entire WKU faculty. Included in the statement were the following goals for WKU-APEP:

- to define each of the four intellectual skills reflect the achievement expected of all students earning a baccalaureate degree
- to specify for each skill area the average skill level attainment expected of students
- to ascertain where in the university, in what courses and experiences, students have the opportunity to acquire each of these intellectual skills

- to determine for each of the four skill areas the average skill level attained by randomly selected students
- to compare expected skill level attainment with actual skill level attainment in each of the four skill areas
- to make recommendations for change to the academic council if the evaluation indicates that student achievement in any skill area is less than that expected by the university
- to prepare a case study of the project experience.

In addition, the statement indicated that the project should result in improved methods for evaluating programs, increased efforts at program evaluation on the WKU campus, and improvement in the effectiveness of instruction.

Implementation

The initial charge of the committee was to define the skills. Conceptual definitions for the intellectual skills were formulated by four subcommittees, each with responsibility for defining a particular skill. Drafts of these definitions were submitted for review to the faculty-at-large and revised by the committee in response to criticism and suggestions. (The definitions that resulted are presented at the end of this chapter.)

Committee meetings continued weekly. Throughout the process of defining the skills, the committee sought the opinions of and guidance from the faculty. Although no attempt was made to hold a forum, interaction between the committee members and the general faculty was encouraged and was found to occur spontaneously within departments, colleges, and the university as a whole.

Assessment Instruments

Achievement of the first goal (the operationalized definitions) also marked the end of the academic year and the completion of the terms of general education committee service for the elected members. To ensure uninterrupted continution of APEP, the project director and coordinator decided to form for each skill a new subcommittee, chaired when possible by an individual who had served on the original committee. Three of the four new subcommittees were chaired by faculty members with no previous involvement in APEP. To obtain a new committee, the college deans were asked to appoint to each subcommittee a faculty member who possessed both interest in the project and the appropriate expertise. Chaired by the WKU APEP coordinator, this committee was charged with reviewing available assessment instruments and selecting or developing instruments suitable for assessing the previously defined generic skills.

Because generic skills development should occur throughout the baccalaureate experience, the degree program should provide an appropriate scope for the assessment of seniors. However, incoming freshmen lack such experience, and a senior-freshmen comparison would not be invalid. Thus, a "life" scope assessment, involving situations that should be equally applicable to both groups, was chosen.

A necessary step before assessment instruments could be selected was identification of subskills, a difficult task in that four of the committee members who formulated the definitions had not been able to continue into this next phase of the project. One advantage of the change, however, was the opportunity to get more faculty members closely involved in the project.

One task of this stage of the APEP was not accomplished: the establishment of performance levels. Like the original APEP committee, the new subcommittees resisted establishment of expected performance levels. Some members believed that expectations of performance were inappropriate in the absence of the actual assessment instruments to be used; also many believed that previous performance data were necessary. Changing from a normative-referenced viewpoint to a criterion-referenced one seemed to be difficult for the latter group.

To accomplish the goal of determining at what points in the university learning experience the four skills are learned, the faculty was surveyed. The results identified the academic areas and the specific courses in which students are given the opportunity to develop the skills defined by the WKU APEP. With goals one and three attained, the groups moved to the assessment task.

The subcommittees decided that a test of two hour's duration for each of the four skills not only would be appropriate for the assessment but also was consistent with the extent of available resources. The main issue facing each subcommittee was the suitability of "home-grown" versus commercial instruments. The identified subskills and the content and scope of the assessment were compared with various commercial instruments to determine the extent of congruence.

The communication subcommittee determined that the written communication portions of COMP/ACT Activities 10, 11, and 12 appropriately addressed the subskills defined for communication.[1] This assessment requires a panel of six raters who must be trained utilizing materials supplied by

COMP/ACT. Analysis by COMP/ACT includes comparisons to norms of performance established by that company.

The quantification subcommittee decided on an assessment instrument composed of multiple-choice items and consisting of three portions--mathematical, statistical, and dimensional. For the mathematical thinking portion of the instrument, the subcommittee selected fifty-three items from an experimental design for assessment of mathematical thinking skills authored by Jonathan Warren of the Educational Testing Service (ETS).[2] The statistical and dimensional portions of the assessment instrument were locally derived. The subcommittee developed item pools, from which twenty-two items for each of the two portions were selected. Thus, the complete instrument for quantification consisted of ninty-seven items.

The analysis and synthesis subcommittees also decided that multiple-choice items would adequately address the skill definitions for analysis and for synthesis. Members of the analysis subcommittee developed a pool of locally derived items and selected twenty of the items from the ETS experimental design. In addition, they elected to purchase the Cornell Test of Critical Thinking, Form X.[3] Members of the synthesis group also developed a pool of locally derived items, from which forty-two items were selected. Additionally, they elected to use twenty-three items from forms 41, 42, 43, and 44 of the ETS experimental design. Thus, sixty-five multiple-choice items constituted the total instrument for measuring the synthesis skill.

Concern for test reliability and validity affected instrument selection and design. Gathering data to establish the reliability and validity of locally developed tests is a time-consuming process, hence the decision to use commercial instruments and portions of the ETS materials as much as possible. Additionally, with the levels of performance expected of WKU students not established, the availability of performance norms was considered to be of great value.

Skill Assessment

A random sample of senior-level students, stratified by college, was selected from among students who (1) were scheduled for May 1981 graduation, (2) had matriculated at WKU as beginning freshmen, and (3) were completing degree requirements in five years or less. These criteria reduced the chances that student gains resulted from experience at another college or university or in a noncollegiate environment such as military service or full-time employment.

Incoming freshmen were chosen from among students who had attended the early orientation and registration sessions held during March and June 1981 in preparation for fall matriculation. Because the ACT scores of these students were available, it was possible to "match" a freshman to each graduating senior. The goal was to test at least fifty students from each population. To encourage the students' cooperation, they were offered free movie and bowling tickets, free meals, and (for the freshmen) free overnight accommodations.

The testing of each group, seniors during spring semester and incoming freshmen during the summer, was conducted over a two-day period with two of the skills evaluated each day. Examinations were supervised by the project coordinator and by subcommittee chairpersons. The four subcommittees worked independently to score the examinations and analyze the data. Abstracts of the subcommittees' reports, prepared by the chairs of those groups, follow.

Synthesis

The final 65-item form of the Synthesis Evaluation Instrument (SEI) was determined largely by the conceptual and operational definitions formulated by the General Education Advisory Committee, chaired by Phillip G. Duff, associate professor in the department of psychology. However, additional criteria for discipline freedom/fairness and face validity were consistently employed in final item selection.

Analyses of data generated by senior and freshmen samples revealed the following:

- *Reliability of the SEI.* When the two reliability coefficients were appropriately averaged, a mean reliability of internal consistency of 0.84 resulted--a value sufficiently high to warrant the adjudgment of significance.
- *Validity of the SEI.* When the correlation coefficients between SEI and ACT performances were appropriately averaged, a mean correlation of 0.67 was derived. The value of this coefficient can be interpreted as evidential support for validity of the SEI while at the same time militating against possible charges of redundancy.
- *Differences between SEI Performance Levels.* The most obvious--and distressing--difference between the performance levels of the two groups related to their respective means. Athough senior sample SEI mean performance exceeded that of the freshman sample, statistical analysis indicated insignificance of that difference.

- *Acceptable Level of Performance.* By utilizing the necessary chance performance statistics, the familiar "z" formula, and a desired level of confidence, a value reflecting statistically significant deviation from change-level performance was identified. This value was then established as the acceptable level of performance on the SEI.

Communication

The Communication Skill Assessment Instrument

The communication subcommittee, chaired by Mary Ellen Miller, associate professor, department of English, selected activities 10, 11, and 12 of the COMP/ACT as the assessment instrument for communication. These activities test for listening comprehension, ability to follow instructions, vocabulary, writing mechanics, organization, development, logic, and persuasion. An obvious advantage of the choice was the availability of norms, which ACT established by using data from seniors at twelve institutions that used Form IV of the COMP/ACT. Even so, the instrument was still considered experimental and under development.

Six faculty members of the WKU Department of English graded the tests. The graded tests were forwarded to ACT, where reviewers judged the grading to be consistent with the ACT scales except for one section, which was graded by ACT staff.

Analysis of Data, Communication Skill

In analyzing the data, ACT staff members coded the senior students by the college in which the major area of study was located and constructed a subsample of freshmen which was matched to the senior group on ACT scores and (as far as possible) on college.

The "average senior" in the small and self-selected subsample of seniors scored at the 39th percentile compared with seniors in the reference group. Compared with the same senior reference group, the freshmen sample scored at the 28th percentile. The matched subgroup of freshmen obtained a mean writing score at the 23rd percentile. Thus, if the samples had been representative, assessment results would have indicated gains of 11-16 percentile points occurred over the period from freshman to senior year.

As might be expected, both seniors and freshmen who reported themselves as being in the college of arts and humanities scored higher than the other college groups. When the seniors were stratified into grade-point average subgroups, no relationship between grade-point average and writing proficiency was evident.

Evaluation of Results and Recommendations

The faculty graders found that the writing performance of some of the seniors was below the level needed to pass the qualifying examination for English 102, required for all freshman students. Since all the seniors tested had presumably passed the qualifying examination, these results suggested that the writing skills of some students had degenerated between English 102 and graduation. The Communications Subcommittee recommended further testing of seniors to confirm or refute that finding. Consequently, the 102 exam was given to one or two classes of seniors in each college, with the grading being done by faculty who normally grade 102 papers. These faculty were not informed as to which papers were written by seniors. In this followup twenty-seven percent of the graduating seniors (compared with 10 percent of the freshmen) were found to have failed the exam. As a result of these findings, a task force on "Teaching Writing and Thinking Across the Curriculum" is now in action at WKU.

Analysis

The Analysis Skill Assessment Instrument

The instrument used consisted of form X of the Cornell Critical Thinking Test (CCT) and a twenty-item test developed by the Analysis Subcommittee, chaired by Larry O. Mayhew, associate professor, department of philosophy and religion) from the analysis scales 31, 32, and 33 provided by Jonathan Warren of ETS. The manual for the CCT suggests that the instrument is a valid test of the following aspects of critical thinking: detecting validity and invalidity of reasoning, locating assumptions, evaluating the reliability of observations and of authorities, generalizing from data, and assessing the relevance of reasons. The ETS scales include items that ask for general interpretation of information presented in a variety of modes (in written form, pictorially, graphically), for locating key assumptions and core issues, and for assessing logical inferences. Because the reliability coefficients of the ETS scales were low, the Analysis Subcommittee adopted eighteen items they deemed to be valid analysis items and completely rewrote two other items.

To gain more evidence of validity (especially for the ETS test), the tests were administered in late fall 1981 to students in an upper-level philosophy course. Before the results were scored, the instructor selected sixteen students whom he judged to be above average (in comparison with the general college population) in analytical ability. (The selection criterion was a grade of "B" or above on two course papers that required analysis of journal articles in the philosophy of mind.) These sixteen students included eight graduate students from the M.A. program in humanities, specializing in philosophy, and eight undergraduate philosphy majors/minors (four seniors, three juniors, and one sophomore). The scores of these sixteen students were used for comparison.

Assuming that this group of students was above average in analytical ability, the results indicate that the instruments selected were able to detect differences in analytical ability. For example, on ETS the freshmen showed a range of scores from 3 to 19, the seniors from 5 to 19, and the (presumed) high group from 15 to 18. On the two instruments combined, the ranges were 34 to 77 for the freshmen, 30 to 83 for the seniors, and 64 to 80 for the high group. (The groups' respective means for CCT+ETS were 56.8, 62.2, and 71.5.) These results indicate that the tests are able to detect differences in analytical ability and provide an encouraging indication of their validity.

Correlation coefficients were computed for the instruments against ACT scores and (in the case of seniors) against grade-point averages. The moderate correlation between grade-point average and the instruments used indicated that the tests measured skills related to but different from academic ability. The moderate correlation between ETS and CCT reinforced intuitive judgments that these instruments emphasized somewhat different analytical skills. The large correlations between scores on the instruments used and ACT scores may have indicated that the skills measured are necessary for good performance on the ACT.

One important question about validity remains. If the group of philosophy students actually did possess a substantially higher-than-average analytical ability, is it possible that this group of students was only 8 percent (or even 16 percent) better than the average freshman at analyzing difficult issues? Obviously, important aspects of analytical ability are missed when only measured subskills are considered. The ability to apply the subskills collectively is probably an important aspect of analytical ability.

Evidence of reliability was obtained by using Cronbach's alpha, as a statistical measure of internal consistency. Reliability coefficients for the ETS were 0.79 and 0.73 and for the CCT were 0.85 and 0.81 for the seniors and freshmen, respectively. Thus, the subcommittee is "cautiously optimistic" about instrument reliability.

On each test, the mean score of the seniors was higher than the mean score of the freshmen, but the difference was not statistically significant. Given the available data, the members of the analysis subcommittee concluded that they were "unable to set standards and evaluate the performance" the seniors.

The setting of performance standards was the problem that seemed "most intractable" to the committee. Without them, no reliable means of interpretation was available. Although the committee considered three different ways of setting standards, serious objections were raised to each procedure. The committee recommended that the freshmen group be retested as seniors to explore changes that actually occur during the college experiences.

Quantification

The Quantification Skill Assessment Instrument

The evaluation instrument developed by the quantification subcommittee, chaired by John P. Russell, associate professor, department of industrial and engineering technology, consisted of three portions identified as mathematical thinking, statistical thinking, and dimensional thinking. The mathematical thinking portion, consisting of fifty-three multiple-choice items, was derived from an experimental design by the ETS. The items selected were those that demonstrated a biserial correlation with the overall test results greater than 0.45 and which fewer than 75 percent of the respondents answered correctly. These criteria were obtained from the results of a trial administration of the ETS design to approximately 1,500 college students at selected institutions. The statistical and dimensional portions of the instrument, each consisting of twenty-two items, were derived from item pools developed by the quantification subcommittee. The items were designed to be similar in scope and difficulty to those of the mathematical items, with the expectation that results from the three portions would be similar. The total instrument, therefore, consisted of ninty-seven multiple-choice items requiring approximately two hours to complete with no special materials required. When the instrument was administered to the group of seniors and to the freshmen, distribution of scores for the three individual portions of the instrument were identical, suggesting that they did provide equal discrimination among the students tested.

The quantification subcommittee chose a minimum performance standard of 50 percent correct responses. The decision was based primarily on the expected performance of college students on the ETS portion of the tests. Mr. Jonathan Warren of ETS had suggested that the "average college student" should score between 55 and 65 percent on the ETS items. Because test results indicate that the three portions of the test were comparable in scope and level of difficulty, the 50-percent level seemed minimally acceptable for seniors.

Of the seniors tested, 23.8 percent did not perform at an acceptable level. The mean score of the seniors was 9 percent higher than the mean score of the incoming freshmen. The committee suggested, however, that "in order to describe a significance for the difference in test scores . . . some measure of sensitivity of the test to demonstrate a difference would be necessary."

When the test scores of the graduating seniors were matched with their composite ACT scores, a positive correlation between these factors was noted. The freshmen were tested in two groups. The fact that the mean test score of the second freshman group was 4 percent lower than the mean score of the first group also suggested such a correlation. Fifty percent of the total freshman group scored below the level deemed minimally acceptable for seniors.

Conclusions

Results of the WKU APEP were inconclusive. The inability to entice all of the selected seniors to participate and failure to establish performance standards for all four skill areas limit the conclusions that can be drawn. Several interesting implications can be made however. In each skill area, comparison of the scores of graduating seniors and the scores of incoming freshmen yielded a difference in mean scores for the two groups, with the seniors' scores being higher. The difference, however, was not significant in every case. Results also suggest that some of the graduating seniors are not performing at an acceptable level in all four skill areas.

The high correlation of the test scores with the students' ACT composite scores would seem to support the validity of the assessment instruments. Assuming that the ACT test measures attributes that are important in acquiring the generic skills, the instruments used in the assessment measure similar attributes or measure the acquisition of such skills (the latter was, of course, the intent).

The ability of the faculty to address cooperatively such broad concepts as the existence and measurement of generic skills without retreating into discipline-dominated stance is impressive. Faculty participants interacted freely and often commented on the benefits of that experience.

Finally, results of the WKU APEP, although inconclusive, provide a basis for further investigation of generic skill attainment. They suggest that the investigation of student outcomes can lead to further questions and that the search for answers to these questions can provide data that contribute significantly to the understanding of the educational profession.

Analysis

Conceptual Definition

Analysis is the process of hypothesizing the nature of the unified whole, identifying and classifying its component parts, determining the relation among parts, and recognizing the organizational principles involved.

Operational Definition

Analysis involves the following abilities:

- to hypothesize the nature of the unified whole
- to identify and classify the component parts so the unified whole
  1. to identify the fundamental elements or component parts
     a. to identify and distinguish among structural elements
     b. to identify relevant details
  2. to classify the essential features
     a. to classify essential facts
     b. to classify implicit assumptions
  3. to elucidate and clarify expressions, statements, concepts, and principles
     a. to infer the creator's purpose, point of view, or traits of thought and feeling expressed in his/her work
     b. to infer the creator's concept of science, philosophy, history, or art
     c. to describe the formal structure and expressive qualities of a work
- to determine the relations among parts of the unified whole
  1. to detect the interrelations among ideas
  2. to identify the information necessary to show cause and effect
  3. to relate historical antecedents to subsequent or contemporary events
  4. to comprehend the interactions among facts, concepts, events, and forces within a conceptual framework
  5. to recognize ways in which method influences results
  6. to identify the effects of motives, personal points of view, self-interests, biases, or prejudices on a communication or source of information

- to recognize the organizational principles involved in the creation of a unified whole
1. to identify the assumptions underlying the framework within which particular data are considered
2. to distinguish among hypotheses, theories, observation statements, normative statements, and definitions
3. to recognize how a theory, process, work of art, or scientific concept unifies, explains, or orders diverse elements.

Synthesis

Conceptual Definition

Synthesis is the process of organizing and integrating ideas, elements, or parts into a coherent whole--a pattern of structure which was not previously apparent.

Operational Definition

Synthesis involves the following abilities:

- to organize ideas, elements, or parts
1. to recognize basic differences, similarities, and other significant relations among items
a. using information and concepts available from current analysis
b. using information and concepts drawn from previous experience
2. to exclude those items not significantly related
3. to recognize possible organizational principles and to choose an appropriate one
4. to order the items in accordance with the organizational principle chosen
- to integrate ideas, elements, or parts
1. to choose and use existing organizational systems and concepts or to develop new combinations or organizational systems when necessary or appropriate
2. to develop generalized expressions as a means of explaining specific current phenomena.

Communication

Conceptual Definition

Communication is the process of conveying information, ideas, feelings, attitudes, and experiences to others and the process of receiving and understanding the information, ideas, feelings, attitudes, and experiences expressed by others, with the recognition that the communication may take a variety of forms.

Operational Definition

Communication involves the following abilities:

- to convey information, ideas, feelings, attitudes, and experiences to others
1. to select materials relevant to a particular problem or limited subject
2. to organize the selected material
3. to present the message clearly, logically, and concisely
4. to observe accepted conventions in the communication process
- to receive and understand information, ideas, feelings, attitudes and experiences
1. to identify the elements of a message received relevant to a particular problem or limited subject
2. to demonstrate an understanding of the message received.

Quantification

Conceptual Definition

Quantification is the process of determining and expressing the measured or relative value of an idea, a concept, or a physical entity.

Operational Definition

Quantification involves the ability to determine the measured or relative value.

- to recognize and understand standards of measurement
1. to know the units of measure appropriate to a particular situation
2. to understand the concept of number versus quantity
3. to choose appropriate entities or concepts to serve as basis for relative measure
- to apply measurement procedures and techniques
1. to recognize techniques applicable to a particular situation
2. to apply physical measures for determining quantity
3. to apply appropriate calculation techniques
- to recognize and apply statistical measures.

Focus and Goals Statement

Each recipient of a baccalaureate degree from Western Kentucky University is expected to have technical or professional competence in one or more fields of study, and each graduate is expected to have had a prescribed amount of academic course work in what is usually known as "General Education." The definition of general education approved within the university expresses the belief that general education is a common academic experience designed to assist students in developing wholesome self-concepts, an appreciation of their fellow man, and an understanding of the breadth and depth of accumulated knowledge. Attempting to ensure that students pursuing baccalaureate degrees will have the opportunity to attain the goals which are inherent in that definition, the university requires each student to complete a specific number of hours selected from approved courses within each of the following categories: organization and communication of ideas, humanities, social and behavioral studies, natural science-mathematics, physical development, and general electives.

Prior to this (1979-80) academic year, no effort has been made at WKU to evaluate the extent to which the meeting of these requirements in addition to "major" (and "minor" for some programs) requirements ensures students' attainment of specific intellectual competencies or of a specific level of knowledge in any one of the stated categories. In September 1979, Western Kentucky University became one of ten institutions chosen to participate in a 2-1/2-year evaluation project developed by the American Association of State Colleges and Universities and partially financed through the Fund for the Improvement of Postsecondary Education. The project's goal is to evaluate student attainment in four basic intellectual skill areas selected by the academic vice presidents of AASCU institutions. The chosen skill areas are communication, quantification, analysis, and synthesis. At Western, the project is directed by James L. Davis, vice president for academic affairs, and coordinated by Faye Robinson, associate vice president for academic affairs.

Since the General Education Committee at Western has the responsibility for continuous purview over general education requirements, that committee (which is composed of faculty and student representatives from the academic units of the university) has been given the charge of working with the project director and coordinator to accomplish the initial tasks involved in evaluating the attainment by Western's baccalaureate students of the specified four intellectual skills. Persons from that committee and from the faculty at large will have extensive involvement as the assessment procedures are planned.

The goals of the Academic Program Evaluation Project at Western Kentucky University are:

- to define each of the four intellectual skills to reflect achievement expected of all students earning a baccalaureate degree
- to specify for each skill area to be evaluated the average skill level attainment expected of students
- to ascertain where in the university--in what courses and experiences--students have the opportunity to acquire each of these intellectual skills
- to determine for each of the four skill areas the average skill level attainment by randomly selected students
- to compare expected skill level attainment with actual skill level attainment in each of the four cognitive skill areas
- to make recommendations for change to the academic council if the evaluation indicates that student achievement in any skill area is at a level lower than that expected by the university
- to prepare a case study of the project experience.

It is expected that this project will result in improved methods for use in program evaluation, increased efforts at program evaluation on this campus and on the other campuses to which project information will be distributed, and improvement in the effectiveness of instruction.

Notes

[1] The American College Testing Program, P.O. Box 168, Iowa City, IA 52243.

[2] Jonathan R. Warren, Senior Research Scientist, Educational Testing Service, Berkeley Office, 1947 Center Street, Berkeley, CA 94711.

[3] Illinois Rational Thinking Project, 188 Educational Building, University of Illinois, Urbana, IL 61801.

Written By

A. Faye Robinson and John P. Russell

Institutional Contacts

Faye Robinson, associate vice president for academic affairs

John P. Russell, associate professor of industrial and engineering technology

Assessing Outcomes: The Real Value-Added is in the Process

Trudy W. Banta and Homer S. Fisher

Across the country, assessment of the outcomes of higher education is a topic of growing concern to faculty and administrators, the various publics they serve, and the agencies that influence their activities. The October 1984 NIE report Involvement in Learning (8) contained the recommendation that each institution assess progress toward attainment of its objectives for student development. Four months later, in Integrity in the College Curriculum, the AAC announced that "The colleges themselves must be held responsible for developing evaluations that the public can respect (1, p. 33)." An October 1985 meeting originally planned to provide a forum for discussing responses to the range of recommendations in the NIE report one year after its publication quickly became the National Conference on Assessment in Higher Education, and attracted over 700 people--nearly double the early estimates of attendance.

U.S. Department of Education Secretary William Bennett has announced to higher education officials that "Those who pay for education must have reliable information about the quality of the institutions they are supporting (6)." The current chairman of the National Governors' Association, Governor Lamar Alexander of Tennessee, has appointed a task force on "college quality" headed by Governor John Ashcroft of Missouri to consider ways of measuring student learning and institutional effectiveness. The Education Commission of the States is studying state strategies, including assessment, for improving undergraduate education. The six regional accrediting associations, which approve private as well as public institutions, have revised their standards to include institutional effectiveness components that require evidence of student accomplishment of educational objectives derived from assessment activities.

Considering the number of individuals, institutions, agencies, and associations studying and attempting to shape outcomes assessment, it is not surprising that differing perspectives on the topic have emerged. An exchange of views at the February 1986 meeting of Governor Ashcroft's task force on college quality illustrates the lack of consensus about purposes, definitions, and methods of assessment. According to Ashcroft, "Assessment leads to quality and accountability. Standards, as measured by tests, send a strong message to young students that they will be held accountable (7, p. 14)." He went on to say that enforcing standards through testing likewise sends a message to professors and administrators that they will be held accountable for their students' performance. Ohio Governor Richard F. Celeste warned, however, that there is "a real danger of going overboard on standardized testing in higher education. The danger is imagining that some simple and uniform formula can be used to say we know how everyone stands (7, p. 14)." Instead of relying solely upon tests, Celeste urged that assessment take the form of thorough peer reviews of departments and colleges that are linked to competitive financing programs.

The comments of the two governors reveal the tension between those--usually outside the institution, in legislatures and state coordinating agencies--who would hold colleges and universities accountable for the quality of educational programs by imposing standards based on criteria that are easy to measure and uniform across a variety of institutions, and those--usually faculty, students, and administrators within the institution-who would prefer to use a variety of assessment activities to determine the status of programs in meeting their objectives and to suggest directions for improvement. While most within and outside academe agree that program quality in higher education should be assessed, there is little agreement about what should be assessed and which methods should be used.

A recent telephone survey of state academic officers conducted by ECS (5) has revealed that all 50 states currently have underway initiatives for improving undergraduate education. Several of these incorporate assessment activities. In their report on the ECS survey, Boyer and McGuinness assert that "State action...is stimulating public colleges and universities to scrutinize their undergraduate programs (5, p. 7)." They also point out that some institutions "are making extensive use of outcomes data in overall institutional planning as well as in curriculum evaluation and improvement (p. 6)." However, they add,

> Limited evidence suggests that colleges and universities can make effective use of outcomes data in their curriculum-reform efforts. Before other states act on that evidence, however, more discussion is needed on how and under what conditions the use of outcomes data can be linked to state policy without accompanying negative consequences (p. 6).

Boyer and McGuinness recognize Tennessee as the first state to provide a financial incentive for public colleges and universities to define their student development objectives more clearly and to undertake assessment procedures to provide evidence of effectiveness in achieving those objectives. From their experience with the Tennessee performance funding initiative the authors propose to contribute to the store of evidence that prudent use of outcomes data can be linked to state policy not only "without negative consequences" but with tangible benefits for institutions.

Performance Funding: Incentive for Assessment

Since 1979 the Tennessee Higher Education Commission (THEC) has been experimenting with a performance funding policy that provides a financial supplement to public institutions for engaging in the following instructional evaluation activities: (1) testing undergraduate students in general education; (2) conducting peer reviews and/or testing undergraduates and candidates for the master's degree in their major field; (3) surveying current students, dropouts, alumni, and/or employers to obtain their assessments of academic program quality and related student services, and (4) using the results of testing and surveys to make needed improvements in programs and services (4). Until 1983 the financial supplement was equivalent to two percent of each institution's budget for instruction; beginning in 1984 the percentage was raised to five. Participation in instructional evaluation activities has been voluntary, but in fact the budget bonus has encouraged every state institution to undertake some or all of these assessment procedures.

The THEC performance funding policy has a number of strengths. First, it was phased in gradually over a period of three years to allow time for institutions to plan their responses to the proposed assessment program. Second, the funds provided represent a supplement to the formula-generated allocation, thus no institution has experienced undue pressure to participate or loss of anticipated funding for failure to do so. Perhaps the most important strength of this state policy is its flexibility: while there are additional rewards for high test scores, institutions also receive credit for undertaking assessment activities and demonstrating that they have been guided by the results in making program improvements. Moreover, the policy does not require institutions to use standardized tests in major fields; each program faculty is encouraged to select or develop a test that measures the achievement of its objectives for students.

Tennessee's performance funding policy thus far has avoided the over-reliance on standardized tests that Ohio Governor Celeste has warned against; it has encouraged the "thorough peer reviews" he advocates; and it has stimulated colleges and universities to use outcomes data in "curriculum-reform efforts" and "overall institutional planning," as Boyer and McGuinness have suggested. The following sections provide concrete evidence of the value added to the educational process and to decision-making throughout the institution by the kind of outcomes assessment program that is suggested by the THEC performance funding policy. Most of the illustrations come from the experience of The University of Tennessee, Knoxville (UTK), the state's land-grant and research university with an enrollment of 25,000, including 5,500 graduate students.

Responding to the Performance Funding Initiative

Performance funding was viewed initially as an external threat by administrators and faculty on many campuses in Tennessee. At UTK some of the negative impressions were at least neutralized in the course of a two-year study carried out with assistance from a small Kellogg Foundation grant administered by NCHEMS. Taking advantage of the time for planning offered by the THEC, three task forces composed of students, faculty, and administrators considered instruments available for testing students in general education and in their major and for surveying students and alumni. Then they made recommendations for integrating outcomes assessment activities in ongoing procedures vital to the life of the institution (2).

A first principle of outcomes assessment is that there is no single set of outcomes, or approach to assessment, that is appropriate for every campus. Each institution must consider its own mission and goals, as well as goals and desired instructional outcomes for individual programs, in deciding what to assess and which of a variety of methods to use. Fortunately, the THEC performance funding guidelines provide some latitude for tailoring an assessment program to address the mission and goals of each institution.

At UTK the Kellogg project task forces enunciated three purposes for a university-wide assessment program:

(1) To establish the status of programs in meeting their objectives for student development;
(2) To provide direction for improving and strengthening curricula, methods of instruction, and related student services;
(3) To improve the quality of information used by departments, colleges, and the university in making decisions about the relative priorities of certain activities and the allocation of institutional resources.

More specifically, the task forces recommended that outcomes assessment be incorporated in the well-established campus peer review process and that the enriched peer reviews be given increased importance in the institution's strategic planning and resource allocation deliberations.

The number of outcomes listed in Figure 1 as options for departments to consider in their self-study prior to peer review indicates that members of the three task forces wanted to go beyond the areas of assessment specified by the THEC. The task force on Measurement of Achievement in General Education asked that the ACT College Outcome Measures Project (COMP) exam be administered centrally to freshmen and seniors to permit an assessment of growth from entry to exit. COMP scores were to be provided periodically to colleges and departments. The task force on Measurement of Achievement in the Major recommended that the faculties of 110 academic programs select a standardized exam or construct their own to assess the status of their largest program (baccalaureate or masters's degree level) in attaining the faculty's objectives for student development in the major field. The task force on Measurement of Satisfaction asked two faculty specialists in survey research to design a series of related instruments to assess satisfaction of students, dropouts, alumni, and employers with the quality of university and departmental programs and services. Oversampling in departments scheduled for peer review would permit the incorporation of survey data collected centrally in departmental self-studies. Beyond the assessment information gathered by the central administration, colleges and departments were encouraged to collect and report their own data on student perceptions of growth toward objectives, placements and achievements of graduates, and external recognition received by the program.

Benefits Resulting from Outcomes Assessment Activities

Since 1982, when a systematic outcomes assessment program was initiated at UTK with findings incorporated in academic program reviews, strategic planning, and resource allocation procedures, benefits have begun to accrue throughout the institution.

Academic Program Review

In 1974 UTK administrators established a peer review process that utilizes a team of two external consultants in the discipline and three internal consultants from units outside the one being reviewed. The team of five reviewers spends two-and-a-half days gathering information from faculty, students, and administrators to supplement that obtained from the departmental self-study prepared in advance of the visit. Prior to 1983, self-study guidelines had encouraged the department to describe its resources, including faculty credentials, abilities of incoming students, size of the library collection, levels of salaries and operating budget, and adequacy of facilities. To accommodate the focus on outcomes assessment, the guidelines were expanded to include suggestions that the department state its instructional objectives for students then provide evidence of its effectiveness in helping students attain the objectives. The addition to the self-study of such evidence as perceptions ofprogram quality obtained from surveys of students and alumni and students' scores on comprehensive exams in general education and the major field has enriched the content of that document. The survey responses and test scores enable the reviewers to focus their questioning of students and faculty on perceived strengths and weaknesses of the program, and many reviewers have expressed appreciation for these new tools for program evaluation.

Strategic Planning and Resource Allocation

At UTK a central planning group of administrators and faculty leaders seeks the best match between internal program strengths and weaknesses and external opportunities and constraints in determining strategy for the institution (3). University - wide surveys of students, alumni, and dropouts have assisted the planning group in assessing both the external and the internal environment of the institution. Students' test scores and departmental survey results, along with other information from the academic program reviews, have helped the planning group determine which programs need additional support to grow and/or to achieve distinction, and which should be reduced in size or scope. Data from outcomes assessment also have been used to set planning goals related to enrollment and marketing; student/faculty/staff development; and improvement of equipment, physical facilities and the library.

While academic program reviews take place only at five- to seven-year intervals, annual budget hearings permit department heads to present to the central administrative staff their goals and objectives and progress toward accomplishing these, and their associated resource needs. With the advent of outcomes assessment activities, many department heads have been able to produce new evidence of both their needs and their accomplishments. At the same time, central administrators have at their disposal improved comparative data for use in setting priorities for the allocation of funds.

Information from the various assessment activities has helped to convince central administrators at UTK that resources should be allocated for the following purposes:

- To strengthen the general education core as the University moves from quarters to a semester calendar.
- To reduce class size in selected areas.
- To focus attention on the need for improved strategies for student retention.
- To establish or expand college advising centers.
- To provide released time for faculty in certain units to participate in advising.
- To improve publications for prospective students.

College and Departmental Decision-Making

The addition of findings from outcomes assessment activities has improved the quality of the information on which deans and department heads base decisions about academic programs and related services such as advising. Administering the ACT COMP exam to freshmen and seniors and observing patterns of growth in the six areas assessed by the instrument has provided information which suggests that the university's general education curriculum, especially in social science, should be strengthened as new patterns of courses are developed for the semester system. Within individual colleges the nature of the general education core has been shaped by the profile of specific strengths and weaknesses of students in that college as revealed by COMP exam subscores.

The locally developed surveys for current students, alumni, and dropouts permit colleges, and departments in the case of current students, to compare their own ratings with those of other colleges and with the university-wide average. Having access to the perceptions of those they seek to serve concerning the quality of academic programs and related services has prompted administrators and faculty in several units to:

- Increase the emphasis on advising.
- Increase opportunities for faculty-student interaction.
- Initiate or expand opportunities for internships that permit students to apply their knowledge and skills in employment settings.
- Improve printed information describing the academic program.
- Initiate or expand student professional organizations.
- Increase faculty involvement in the placement of graduates.

Both deans and department heads have found that alumni, employers, advisory council members, and potential donors to their programs are favorably impressed by the seriousness with which their units have approached the collection and use of outcomes information. The ability to demonstrate high levels of student and alumni satisfaction has given a boost to development efforts across the campus.

As faculty have become involved in the development of comprehensive exams for majors, they have evolved fresh perspectives on the nature of the curriculum and the knowledge and skills students should derive from it. The test development process itself has produced a number of positive outcomes at the departmental level, including:

- Increased faculty agreement concerning common learning outcomes for all students.
- Development of core objectives or competencies for students.
- Increased direction for converting courses/curricula from quarter to semester format.
- Increased consistency in the teaching of core courses.
- More clearly delineated linkage between lower-division and upper-division courses.

Following administration of the locally developed comprehensive exams and joint review of students' performance, faculties have:

- Acquired baseline data on student achievement that can be used to compare the effectiveness of quarter and semester curricula.
- Identified program strengths, and weaknesses that should be corrected in the semester curriculum.
- Made curriculum changes, including
  - ·Stronger core curriculum within the department, and
  - ·Changes in requirements outside the department.

Faculty

As a result of looking at students' ACT COMP scores on the Communicating, Solving Problems, and Clarifying Values subscales, and reviewing students' scores on comprehensive exams in major fields, faculty have made a number of modifications in their own instructional techniques, including increases in:

- Structure for courses now that core objectives/competencies are identified.
- Number of written assignments.
- Opportunities for students to apply their knowledge and skills through problem-solving, term projects, field trips, internships.
- Opportunities for demonstrating complex skills on tests and in other assignments, and
- Use of scores on the local test to diagnose students' needs for remediation prior to taking the certification exam in the discipline.

Instructors increasingly are developing improved and more specific objectives for student learning and relating course content to those objectives, then testing students to ascertain their accomplishment of the objectives. Their test development experience has placed UTK faculty in a position to be called upon to set standards and develop examinations for use at the national level by their respective professional associations. Several faculty involved in test development have produced papers and/or articles for publication in journals dealing with instruction in their discipline. Obviously these activities have enhanced the qualifications of some faculty for promotion and increases in compensation.

Participation in assessment has pointed out the need for certain types of faculty development activities on the campus. Faculty have shown increased interest in such topics as writing items for classroom tests, improving students' writing abilities, and increasing students' abilities to engage in critical thinking and problem-solving.

Students

Students at UTK have benefited in a number of ways from the institution's instructional evaluation activities. As faculty expectations for student development have been clarified and communicated more effectively, students have been able to bring their own educational goals into sharper focus and improve the efficiency of their learning activities. Students' academic satisfaction was increased as faculty have attached added importance to good advising and interacting with students outside the classroom. Finally, most students have welcomed the additional information about their growth toward mastery of curricular objectives that has been afforded by the comprehensive exams in general education and the major.

Summary and Conclusions

Consensus is growing concerning the need for assessing levels of student learning and the quality of the educational process in our colleges and universities. However, there is no single set of procedures that will permit every institution to assess its student outcomes appropriately, nor does the current state of the art of measurement offer a set of procedures that will permit any institution to assess all of the intended outcomes of the education it provides. Nevertheless, there is a growing body of evidence that a given institution can select and/or develop procedures for outcomes assessment that support its mission and goals, and can use the results of these procedures to effect improvements in programs and services that will benefit students, faculty, and administrators, departments, colleges, and the institution itself.

Assessing educational outcomes necessarily involves using tests and questionnaires that elicit relatively standardized responses and thus yield scores or ratings that are easy to interpret. Faculty and administrators can and do raise legitimate questions about the technical qualities of these "objective" instruments. However, when students' responses are not used to make decisions about individuals, but rather studied in the aggregate to gather clues about the strengths and weaknesses of programs and student services, some deviation from technical perfection may be tolerated.

Program quality must not be judged exclusively on the basis of factors that are relatively easy to quantify, such as student achievement in cognitive areas and client satisfaction. Over-reliance upon the absolute value of comprehensive test scores alone, without regard to the amount of student growth inherent in those scores, and without related information about the educational process that helped to produce the scores, will strengthen the critics of assessment--those who maintain that we should not attempt to assess any outcome until reliable and valid measures exist for assessing all outcomes of higher education.

Levels of scores by themselves can be used to establish the status of a program vis-a-vis its own goals or those of an external agency. But this is a static process. The key to an effective outcomes assessment effort is to initiate a dynamic process--using multiple sources of data to suggest program strengths and weaknesses and potential directions for improvement, then relying upon human judgment to process outcome data in combination with information about program resources in formulating recommendations for change.

More than six years of experience indicates that the current performance funding policy of the Tennessee Higher Education Commission is sufficiently flexible to permit colleges and universities to look beyond the simplistic information provided by levels of test scores to seek directions for improvements from multiple data sources. The financial carrot offered by performance funding has stimulated faculty and administrators at institutions in Tennessee to (1) formulate more specific objectives for student development, (2) communicate these objectives to students, (3) involve students in learning through increased faculty-student interaction and opportunities to apply knowledge and skills acquired in the classroom, (4) assess student progress in attaining objectives for their development, (5) assess student satisfaction with their progress and with the process used to promote it, and (6) make improvements indicated by the assessment data. It is in this dynamic process of using outcomes information that the real value added by higher education will be ascertained and affirmed.

References

Association of American Colleges. The Problem of Accountability." In Integrity in the College Curriculum: A Report to the Academic Community. Washington, D.C.: Association of American Colleges, 1985.

Banta, T.W. Performance Funding in Higher Education: A Critical Analysis of Tennessee's Experience. Boulder, CO: National Center for Higher Education Management Systems, 1986.

Banta, T.W. and Fisher, H.S. "Performance Funding: Tennessee's Experiment." In J. Folger (ed.) Financial Incentives for Academic Quality. New Directions for Higher Education, no. 48. San Francisco: Jossey-Bass, 1984, 29-41.

Bogue, E.G. and Brown, W. "Performance Incentives for State Colleges." Harvard Business Review, 1982, 60 (6), 123-128.

Boyer, C.M. and McGuinness, A.C. "State Initiatives to Improve Undergraduate Education, ECS Survey Highlights." AAHE Bulletin, 38 (6), February 1986, 3-7.

"Colleges must rate themselves, Bennett says." Knoxville News-Sentinel, Thursday, October 24, 1985, A-5.

Jaschik, S. "Fear of Federal Cuts Tempers Governors' Zeal for Education." The Chronicle of Higher Education, 32 (1), March 5, 1986, 11, 14, 15.

Study Group on the Conditions of Excellence in American Higher Education. Involvement in Learning: Realizing the Potential of American Higher Education. Washington, D.C.: National Institute of Education, 1984.

Trudy W. Banta is Research Professor, Learning Research Center and Homer S. Fisher is Executive Vice Chancellor, at the University of Tennessee, Knoxville, 1819 Andy Holt Avenue, Knoxville, TN 37996-4350

This paper was published in part by the authors in "Putting a Premium on Results." Educational Record, 67, 54-58.

Figure 1

OUTCOMES

Achievement in general education (via COMP exam)
Achievement in major (national or local tests)
Perceptions of growth toward objectives
Opinion of program quality (students, graduates, dropouts)
Opinion of employers concerning job-related competence of graduates
Rate of job placement for graduates
Rate and quality of placement in graduate or professional education
External recognition (honors, awards) achieved by students or graduates
National program rankings

ACADEMIC PROGRAM REVIEW

PROGRAM IMPROVEMENT

STRATEGIC PLANNING

RESOURCE ALLOCATION

ASSESSMENT AND THE REIFICATION OF LEARNING

Clark Bouton

In this paper, I would like to look at assessment from the point of view of the classroom teacher. While I have had some experience from the other end, I am now back to full-time teaching. Both my research and practical efforts are now focused on understanding how learning occurs in the classroom and the interactions which encourage it, inhibit it or determine what kind of learning will occur.

What assessment is assessing is of course what happens in my classroom and hundreds of others. No matter how grand the scale of higher education, it is only comprised of multiples of these classes. Learning happens there or it doesn't happen. Rather than talk about how assessment is done, I would like to talk about what is being assessed. We have, I think, concentrated too much on how and too little on what we are trying to assess. My paper will in a sense be about assessing the assessors.

I picked up the announcement of this Conference one day as I was returning from a class. I was still trying to sort out what had happened in the class session with this small group of students. As I observe my students, I see all kinds of learning taking place--some planned, some accidental, some involving deep insight in a student, some dull and superficial, some learning which will endure and affect the students' lives, and some which might reappear on an examination but then be surely forgotten. Sometimes little or no learning takes place that I can see, and at other times I can't keep up with it. A lot of the most important learning is of the sort that I would not know how to test for; I know it occurs because I experience it happening. A student suddenly makes connections which illuminate aspects of her own life experience as well as the subject matter at hand. Another student finds real power in a concept that most of the students find hardly interesting, and I realize that it has become part of his way of seeing the world. In the same way, I sometimes notice a student is practicing skills and abilities with new confidence, greater willingness to risk, or increasing effectiveness.

As I read the announcement of the Conference and reflected on my class, I was struck by the distance between by students' learning and the assessment of "outcomes" to be discussed here--of the tenuous relation between the two. As I struggle to understand the learning in my small class I wonder at the hubris of those who assess whole programs and institutions. There must surely be some fundamental difference in the way we conceive of the task--not merely of how, but of what is being assessed.

When I look from my small class, I see a vast industry for educating students. I hear a language that speaks of "providing education," of students "getting an education," of knowledge being "transmitted" to students, and even of "alternative delivery systems." Knowledge is quantified into standard "units" and "credits," ratios" and "equivalents." And finally there is talk of "accountability" and assessment."

We have grown used to this language and scarcely hear the conception of education embedded in it. However, what we take for granted represents a radical development in the meaning of education. From being an activity of the learner, involving the change and growth of the person in the process, education becomes reified. It becomes a "thing"--something with a reality external to the learner, something to be "acquired."

The direction of higher education has been towards an ever greater rationalization of learning. The complex, organic learning process has been broken down into the smallest component parts--specific skills and subject matters. These parts are supplied by increasingly specialized teachers, dealing with progressively smaller and more abstract pieces of education in large bureaucratic institutions. Within this extreme division of labor and fragmentation of learning, neither teacher nor student are able to experience education as an integral process of development.

The content of learning is standardized into measurable comparable units, defined in terms of time. Time loses its qualitative, variable nature, relative to the particular activity. It becomes a quantifiable continuum, broken into standard "periods" into which learning must fit, and everything is eliminated which cannot fit within that rigid structure.

The system can include only what is rational, in the sense of calculable and predictable. This has meant the progressive elimination of the qualitative, human and individual attributes of the student and teacher. Almost all learning is fitted into the course structure, and both the content and process in the courses are prescribed. Interaction in the classroom is restricted to impersonal and unemotional communications narrowly focused on the academic task. The content must be "equivalent" to that of other academic courses carrying the same label. What falls outside this standardization may be tolerated, but it cannot be fitted into the rationalized system of education.

Within this rationalized system, knowledge exists in the abstract. It is freed both from the personal relation with teacher and student and from the later use the student will make of it. It acquires the appearance of an external and independently existing thing. The task of the teacher is to transmit to the students a small defined piece of that knowledge, and the task of the student is to acquire it.

This system has effected a profound change in the meaning of education. What begins as the goal of the student to educate himself with the assistance of teacher and fellow students, becomes the goal and responsibility of the teacher to educate the student. The student is alienated from his own learning. Education is experienced as something external to him rather than as his own accomplishment. The teacher decides what the students will learn and how they will learn it, assumes responsibility for seeing that students learn, and uses a system of external rewards and punishments to see that they do. The students also come to see the teacher as responsible for their learning, even to the point of complaining that some teachers, "do not make us work hard enough." What began as their goal and their activity begins to be perceived as something done to them.

Now perhaps we can understand why I find it so difficult to comprehend the learning of my few closely observed students, while others are able to confidently assess tha learning of students on the minimal information provided by a few standard questions. The problem is that I am too close, that I have too much information; I have interacted with them in too many ways. I know them as distinct personalities, pursuing only similar education goals in their own unique ways. They have learned some of the same things, but each has learned these in his own way, has given them his own meaning, and will use--or not use--this learning differently. From where I sit, to standardize the process would lose what I see as most valuable--the differences, the unique qualities of each. I am unable not to see each student as pursuing his own education, starting from a particular place, moving with his own ability and style, and looking forward to his own goals. My task is to assist each of them. From this perspective, invidious comparisons between them seem largely uninteresting and irrelevant.

The situation is far different when learning has been reified. Assessment follows naturally. There is now something objective, standardized and quantifiable to measure. It is no longer the personal goal and activity of the student, but a commodity provided by teachers and institutions. Knowledge is transmitted to the student, and the student can be tested to see whether it took. Hermann Broch described such a teacher as a banker making small loans to the students through the course of the school term, to be paid in full at the end, principal and interest. Given the standardized nature of the knowledge, the equivalent procedures from one course to another, and the objective form of the testing, one need only measure outcomes. The testing can take the form of an input-output model in which the student is treated as the "black box." One need know nothing about the student or the learning process. Assuming these factors are standard, the student can be made to bear the penalty if the specified learning does not occur.

In the same way one may assess the knowledge delivery system. As the focus has shifted from student learning to teaching, by assuming standard goals and students of average ability, the teacher can be evaluated on his efficiency in "getting across the course content." Equally, assuming standard goals and measurements, alternative teaching methods can be compared as to their efficiency. Finally, we reach the level of assessment proper, in which curricula, programs and institutions are evaluated on their ability to deliver the goods.

Why is it that the system of higher education seems ready made for this assessment? The abstraction and reification of knowledge, the standardization and quantification which make assessment possible are already built into the system. Is this merely an accident? Or rather are these not two aspects of the same process of the rationalization of education.

Why did this rationalization take place? Why was this educational model and these methods adopted? Surely not because they are superior methods for encouraging learning. The predominant lecture method, the emphasis on the learning of factual information, the focus on testing and grades have all ben seriously called into question by the research results of the last thirty years. But the research findings have had minimal effect on educational practice. In my own study of college classrooms, I have often found that the procedures used were those least capable of promoting learning. Many of the reasons for this go beyond our concerns here. But the experience has taught me to look for reasons other than the imputed effectiveness of the conventional methods.

The conventional educational methods in higher education seem to have been chosen not because they maximize student learning, but because they correspond to the rationalization I have been describing. They may not promote much learning, but that learning is standard, testable and predictable. It provides a mechanism for sorting out people on some basis of merit. And it can be easily organized and controlled by the administrators. The demands of testing determine the kind of educational methods used, rather than the evaluation process adapting to whatever educational methods seem to best promote learning.

Testing, grading and credentialing play such a dominant role and are so taken for granted that we seem not to notice that educational programs can and do successfully operate without any of these. We seem not to imagine the possibility that people might enter college to meet their individual learning objectives, do the amount of study they are motivated to do, and leave when they've met their objectives--that our task might be to assist the students in meeting these objectives, and that the responsibility for their learning is their own.

Assessment does not measure education without affecting it. By focusing on fragmented, measurable units and ignoring the more complex aspects of the learning process, assessment tends to reinforce educational practices based on the same reified conceptions.

Let me give an example of this. A few years ago I was directing a faculty training project and doing research on student learning groups. In reviewing the literature comparing the effectiveness of group learning methods to the traditional teacher-focused methods, I was first encouraged to find that the group approach generally proved to be more effective, but was then dismayed to realize that in the tests of effectiveness the real advantages of the group approach were not, and could not, be taken into account. The researchers regularly fell back on traditional measurements of outcomes--primarily objective tests on factual content. Some did try to discriminate in a limited way between factual knowledge and some measure of what the student could actually do with the information. But many of the things we saw students doing in the groups, things which we thought promoted the greatest learning and developments simply couldn't occur in the traditional methods. The students' increased independence and responsibility, their practice in expressing their ideas, in giving and receiving criticism, in working effectively with others, their increased confidence and selfesteem--these did not fit into the research. While validating the group method, the research missed what was really valuable in it.

There is a limit to the ability of the assessment procedures to measure the more subtle and complex aspects of learning. And the more influence such measuring exerts, the less inclined will teachers be to devote time and effort to "what doesn't show." In consulting with other teachers, I have often experienced their anxiety over classes in which they actually felt very sure the students were learning more than in their traditional classes, but they were obsessed with "covering the material" and having students show up well on the examinations. They became like the students: if it doesn't show up on the test it doesn't really count.

Too often, assessment is only evaluating what it has itself created. If certain outcomes are tested for, they will be taught for, and teaching and testing are caught up in a cycle of mutual validation. What begins as a modest effort to get a roughly approximate measure of knowledge or ability by using an instrument of some barely understood relation to what it presumed to measure ends by the measure becoming the definition of the thing itself. I need only mention the way IQ score became synonymous with intelligence. I often see this with students who evaluate their performance on their grade in a course when the grade may be the least reliable and informative basis on which to evaluate their performance.

The assessment of outcomes is a relatively easy and simple process situated between the complex learning process and the equally complex employment of knowledge and skills in the real world. As such there is a persistent temptation to see it as an equivalent and substitute it for those processes which are much harder to understand and evaluate. Research over the years has been unable to find a correlation between success in college and success in a profession, even when the education is directly aimed at preparing students for the profession. The knowledge and abilities measured by the testing and grading process seem to have no discernable relation to those needed outside college. Yet this finding seems to have exercised little cautionary effect on our fascination with testing and assessment.

While there is much activity in testing and assessment, there is very little study of the actual process of learning in higher education. It is much easier to devise outcomes measurements than to understand the complex learning process in the classroom. But unless assessment is based on the experience of what happens in the classroom, it is limited to selecting certain outcomes and validating those educational methods which produce the outcomes as defined by the assessment instrument. It cannot know whether these are the most important things happening in the classroom, or whether learning can occur there that is not included in its assessments. If we are to understand what students are learning in college classes, and can learn there, we need first of all to know what they are doing in those classes. We need especially to look at classes that do not conform to the rational model taken for granted in the usual assessments.

Clark Bouton is Professor of Sociology at the University of the District of Columbia, 4200 Connecticut Avenue, N.W., Washington, D.C. 20008

THE RELATIONSHIP BETWEEN HIGH SCHOOL WRITING EXPERIENCES AND UNIVERSITY ADMISSION REQUIREMENTS

John E. Creeden

Introduction

In the past decade, an increasing number of colleges and universities have become concerned with the academic preparation of newly enrolled students. Questions about the deficiencies of college-bound students in mathematics and verbal skills are asked with alarming frequence by faculty, staff, and administrators. Data collected from national studies of students' writing and quantititative skills indicate that there has been a serious erosion of proficiency in these areas. The downward trend in SAT and ACT scores for more than the past decade has added support to the contention that students are not adequately prepared to begin college-level study (Biemiller, 1981; Ferguson, 1978).

At a time when enrollment declines threaten the financial stability of most colleges and universities, an institution's decision to deny admission to an increased percentage of the applicant pool because students are poorly prepared would have a significant impact on the undergraduate population on that campus. As a result of the student activism of the 1970s, colleges and universities have recognized that they have an important role to play in maintaining equal access to education for previously deprived segments of society. Higher education has a broader social responsibility now than in the 1950s and 1960s when a smaller portion of the population was expected to go on to college.

Educational leaders are confronted with a most difficult problem. As more students enroll with serious academic deficiencies, more money and instructional staff time are devoted to teaching remedial courses. In 1979-80, the University of California System estimated that it spent $5.5 million for remedial programs in Mathematics and English (McCurdy, 1981). The University of Wisconsin System conservatively estimates that it spends $2.5 million in direct costs for remedial purposes (Lyall, 1982). If universities continue to support remedial courses, these expenditures will detract significantly from fiscal resources that are needed to fund advanced level courses in regular academic departments. And yet if colleges and universities fail to provide educational opportunites to certified high school graduates who are said to lack basic skills, questions about higher education's social responsibility for educating the public can be legitimately raised.

Recently, a number of studies on the national level have documented the actual decline in the academic skills of high school students (Goodlad, 1984; Boyer, 1983; National Commission on Excellence in Education, 1983). In an effort to communicate to college-bound students the importance of completing a sound high school curriculum, colleges and universities have begun to raise admissions standards (Scully, 1981) with the hope that high school students will improve their preparation for college study. There are serious questions, however, about whether increasing admissions requirements actually solves the problem of inadequate preparaton at its most fundamental level. Many of the students who are admitted to major universities apply with more than the minimum number of required high school courses, yet these are the students who are enrolling in remedial mathematics and English courses in increasing numbers. Simply increasing the admissions requirements does not address this problem adequately.

Writing As A Process

A review of the research literature reveals that the way to improve the writing skills of students is not to apply a return-to-the-basics approach by merely adding more courses in English and grammar instruction. Writing must be seen as a process with a number of interlocking steps through which the writer proceeds, returns, passes through again, and eventually leaves behind as he advances to a more adequate forumlation of his ideas (Emig, 1971; Stallard, 1974; Pianko, 1979). Prewriting, writing, and rewriting are not linear steps in a process, once completed never to be experienced again. Good writing is recursive, and good writers proceed by retracing their steps and reformulating their ideas (Flower & Hayes, 1981; Murray, 1978).

Methodology

The purpose of this study was to look in a more detailed fashion at the high school preparation of freshmen in the College of Letters and Science at the University of Wisconsin-Madison. More specifically, the study was designed to focus on the high school writing experiences of this group of students and determine if there was an association between the amount, frequency, and type of high school writing experiences and an individual student's score on the University of Wisconsin English Placement Test (UWEPT), a 100 item objective test designed to measure editing skills that are considered prerequisites to effective writing. The entire test is machine scored and has a standard score range of 150-850. The placement test, which is administered to all freshment at UW-Madison, is used to determine if students need remedial writing courses or if they are elegible for honors-level English literature courses.

The researcher developed a questionnaire that collected information on students' high school writing experiences. The variables used in the construction of the questionnaire were drawn from the recent research conducted on composition instruction. The questionnaire was modified after it was intially administered to a selected group of students and reviewed by faculty members in the Departments of English and Educational Administration at UW-Madison. For this study, a random sample of 65 students was drawn from the entering freshman class of 1983 in the College of Letters and Science. Each student was personally interviewed and asked questions about his use of the good writing techniques documented in the research literature. Pearson product moment correlations were used to determine if there was an association between the independent variables describing high school writing experiences and an individual student's score on UWEPT.

It was the intent of this study to identify those types of high school writing experiences that contributed to high scores on the English placement test. The findings of the study provide an alternative for those who are attempting to resolve the problem of poor writing skills by raising admissions requirements. Rather than increasing the number of required courses, educators can use the findings of this study to change the content of their high school curricula so that techniques important to the development of good writing skills can be emphasized.

FINDINGS

Characteristics of Sample Students

As a group, this sample of students was similar in many respects to the students in the entering freshmen class, especially in terms of high school rank, academic units completed in high school, and the percentage of students who took English 101, the remedial English class at UW-Madison. Like the students in the freshmen class, this group of students completed far more than the minimum 12 academic units of high school courses required for admission.

In the entering class of 1983, the average number of high school academic units completed was 17.1. For students in the sample, the average number of units completed was 17.7. The average high school rank for the freshmen class was 80.6%; for the sample it was 84.4%. Approximately 9.7% of the freshment class was required to enroll in remedial English 101 compared to 9.2% for the sample group. The average UWEPT score for the freshmen class was 586, while the average for the sample group was 619. More than half (61%) of the students in the study completed 4 years of English in high schol, the maximum preparation, while 57.8% of the freshmen class did the same. The remainder of the students in both the sample and the entering class of 1983 had at least three years of high school English.

Based on the large number of academic units completed, the impressive high school ranks, high UWEPT scores, and the maximum preparation of four units (years) of high school English, one could conclude that these sample students were well-prepared to begin their college study. Such a conclusion, however, does not survive scrutiny. Readiness to begin college-level writing cannot be determined on the basis of number of courses alone. In the next section we will see that this seemingly well-prepared group of students had done little writing in English classes and hardly any paragraph length writing outside their composition classes in high school. Descriptive data available at the time of admission do not indicate the extent of high school preparation or the level of readienss to begin college level course work.

Writing in Composition Courses

It is important for students to select composition courses in high school to help them develop good writing skills (Bamberg, 1978). Almost half of the students in the sample reported that they had enrolled in one or more composition courses while in high school When in composition classes, most students were writing paragraph-length material 6-8 times per semester and were regularly revising their drafts before completing final copies for evaluation, two practices that are found in the literature to characterize good writers (Britton, 1975; Murray, 1978). Among those students who took composition classes, 60% indicated that they regularly showed their work to others while they were revising. As the literature review demonstrates (Bruffee, 1978; Hawkins, 1980), peer review is an important technique that has been effectively used in developing good composition skills. Based on the findings of this study, the majority of sample students did receive formal instruction in composition and were being taught important techniques of good writing.

Writing in Literature Courses

Students reported limited writing opportunities in English courses that were not devoted exclusively to composition. More than half of the students (61.6%) reported that they wrote only one assignment per month of paragraph length material in their non-composition English classes. When writing for literature classes, 60% of the students reported that they regularly revised their rough drafts. More than half of the sample students (53%) reported that they "often" asked others to review their drafts before preparing a final copy.

Both of these practices of revision are important techniques used by good writers. However, there was no difference in the use of revision techniques in literature courses between those who took composition courses and those who did not. This raises serious questions about how much of the instruction in composition students are applying to other writing opportunities. If students are revising in composition courses only because it is expected of them, then composition instructors are not doing a good job of teaching students that revision is an important and necessary part of good writing (Shaughnessy, 1977) and not simply something that has to be done to satisfy a composition teacher's expectations. Although we cannot establish why students who took composition courses did not use techniques of revision in other English Department classes, it is not surprising given Applebee's (1981) findings that only 59% of the English teachers he surveyed nationally reported regularly asking students to rewrite their material. When in non-composition English classes, students in this sample reported limited writing opportunities and did not employ regularly the techniques of revision or peer review.

Writing in Non-English Courses

Based on this study, sample students did very little writing outside the English Department classroom. Approximately 75% of the students reported that they had six or fewer non-English classes in high school where they were asked to write paragraph-length material. When they were enrolled in classes that required writing, in the majority of cases the writing occurred only one to three times per semester. The use of self-revision or revision by others was even more limited in non-English Department classes than in literature classes. And once again there was not evidence to suggest that those who took composition courses applied these techniques more frequently than those students who never took any composition classes. Students in the sample clearly did not take many classes outside the English Department where they had the opportunity to practice any writing skills they may have developed.

This conclusion is similar to those reached by Boyer (1983), Goodlad (1984), and Applebee (1981) about the rather limited writing opportunities on the secondary school level. The research literature indicates that simply increasing the frequency of writing alone does not improve writing skills (McColly & Remstad, 1963; Wolf, 1966).

Use of Prewriting Activities

The research literature demonstrates the importance of prewriting activities (Perl, 1979; Britton et al., 1975; Shaughnessy, 1977; Flower & Hayes, 1981). When asked about the frequency of their prewriting activity in all classes, 72% of the students reported that they "often" spent time organizing their ideas before they began to write. It is important, however, to remember Applebee's (1981) finding that the time span for prewriting activities was rather short, consisting on the average of just over three minutes. When asked what type of prewriting activities they most frequently used, 50% of the students in this study indicated that they simply "jotted down ideas" before they began to write. Although there is no evidence to suggest that any one type of system is the most effective method of prewriting activity, the literature did suggest that careful planning and reflective pausing were crucial parts of the writing process. It is highly doubtful that this type of quality prewriting can occur in only three minutes, and it certainly includes more than "jotting down ideas." Sample students did not engage in the type of deliberate planning and pausing that is characteristic of good writers (Britton et al., 1975).

Relationship Between Variables of Interest and UWEPT Score

The objective of this study was to measure the strength of association between the variables describing high school writing and an individual student's score on UWEPT. Pearson product moment correlations were used to test for the level of association. Only three variables that describe high school writing experiences showed a significant positive correlation with UWEPT score when Pearson product moment correlations were completed (p .05). The variables are: the number of composition courses completed (r = .29), frequency of the use of self-revision in composition courses (r - .26), and frequency of prewriting activities (r - .20). The variables developed to measure the frequency of writing and the use of revision in both literature and non-English Department classes where writing was required showed no significant correlation with UWEPT score.

The findings indicate that instruction in composition does make a difference in a student's score on UWEPT. The more composition courses a student completes, the better the likelihood for a higher UWEPT score. This conclusion is similar to that presented by Bamberg (1978). It makes sense that there is an association between a test such as UWEPT, which measures editing skills considered essential to good writing, and the number of high school composition courses completed by students, where they would presumably learn those skills. A significant difference in UWEPT score was discovered only for those who took two or more composition courses. There was no significant difference in score between those who took one or no composition courses in high school.

Students who use revision in composition courses and those who engaged in prewriting activities tended to have higher UWEPT scores than those who did not employ either technique. Again the importance of the instruction in composition, where the use of revision and prewriting activities would be taught, is essential. Murray (1978), Britton et al. (1975), Shaughnessy (1977), and Huff (1983) have all demonstrated that revision is a fundamental component of the writing process of any good writer. Prewriting is a critical stage in the writing process that allows good writers to explore ideas and discover creative possibilities for structuring their written material (Perl, 1979; Odell, 1974; Elbow, 1973).

UWEPT does appear to be able to detect differences in the particular high school preparation a student has had in written composition. This is encouraging since a recent Wisconsin task force on high school and college collaboration (Joint Council on College Preparation, 1983) suggested that college-bound students include in their high school curriculum at least one course in composition. Based on this study, however, it appears that one course will not have a significant effect on UWEPT score and, therefore, students might want to consider enrolling in more than one composition course.

An association was not discovered between UWEPT score and the variables developed to measure frequency of writing and use of revision in both literature and non-English Department courses where writing was required. This finding is especially troublesome. The data from this study confirm the conclusions of Applebee (1981), Boyer (1983), and Goodlad (1984) that writing is infrequently required at the high school level. The writing in literature and non-English Department classes that does occur, and the use of specific good writing techniques in those classes are not detected, however, by UWEPT. The data do not show a tendency for students who wrote frequently and revised regularly in literature and non-English Department classes to have higher UWEPT scores than students who did not engage in these practices. UWEPT does not appear to be capable of measuring differences in the extent of students' writing across the curriculum. If University administrators believe that curriculum planners will soon follow the recommendations of important studies such as A Nation At Risk (1983) and make writing an integral part of all academic disciplines in high school, then changes will need to be made in English placement tests at UW-Madison so that differences in students' writing across the curriculum can be measured.

Several writing-related variables showed strong association with UWEPT score when Pearson product moment correlations were completed (p .05). Total high school academic units (r = .38) and high school rank (r = .47) were strongly correlated with UWEPT score, although it is difficult to identify with any degree of certainty the specific reasons for the association. Remember that as a group, this sample of students was well-prepared in terms of academic units and high school rank. As a test, UWEPT appears to be better at confirming overall high school performance, which is generally reflected in the combination of total academic units and high school rank, than in specifically measuring the extent of good writing skills developed by students while in high school.

Conclusions

It was clear from the findings of this study that in its current form, the University of Wisconsin Placement Test only measures some of the important techniques that are components of good writing. The University's ability to measure with any sophistication the development of a range of important writing techniques among entering freshmen is severely limited because of the narrow focus of UWEPT on assessing editing skills.

State-wide commissions. authors of national studies on the decline of academic skills, and those who determine entrance requirements must realize the simply raising admission standards does not solve the problem of poor writing among college-bound student. Adding a fourth year of English to the college-preparatory curriculum at the high school level will not by itself foster good writing. Students must be taught and regularly employ the techniques of prewriting, writing, and revision across the curriculum if they are to develop good writing skills.

REFERENCES

Applebee, A. N. (1981). Writing in the secondary school, English and the content areas. Urbana, IL. National Council of Teachers of English.

Bamberg, B. (1978). Composition instruction does make a difference: A comparison of the high school preparation of college freshmen in regular and remedial English classes. Research in the Teaching of English=, 12(1), 47-59.

Biemiller, L. (1981, October 7). 18 year decline in aptitude test scores halted this year, College Board reports. The Chronicle of High Education, 23, 1.

Boyer, E. L. (1983). High school, A report on secondary education in America. New York: Harper & Row.

Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975). The development of writing abilities (11-18). London: MacMillan Education LTD, A report from the Schools Council Project on Written Language of 11-18 Year Olds, based at the University of London Institute of Education, 1966-1971.

Bruffee, K. A. (1978). The Brooklyn plan: Attaining intellectual growth through peer-group tutoring. Liberal Education, (64), 447-468.

Elbow, P. (1973). Writing without teachers. New York: Oxford University Press.

Emig, J. (1971). The composing process of twelfth graders (Research Report # 13). Urbana, IL: National Council of Teachers of English.

Ferguson, R. (1981). The decline in ACT test scores: What does it mean? Educational Technology, 16, 21-27.

Flower, L., & Hayes, J. R. (1981, October). The pregnant pause: An inquiry into the nature of planning. Research in the Teaching of English, 15(3), 229-243.

Goodlad, J. I. (1984). A place called school, prospects for the future. New York: McGraw-Hill Book Company.

Hawkins, T. (1980, September). Intimacy and audience: The relationship between revision and the social dimension of peer tutoring. College English, 42(1), 64-68.

Huff, R. K. (1983, December). Teaching revision: A model of the drafting process. College English, 45(8), 817-823.

Joint Council on College Preparation. (1983). Preparation for college: A statement prepared to assist college-bound secondary school students in Wisconsin. Madison, WI: University of Wisconsin System.

Lyall, K. C. (1982, December 9). The competency and proficiency program of the UW system: Statement to the education committee, University of Wisconsin Board of Regents. Madison, WI: University of Wisconsin System.

McColly, W., & Remstad, R. (1963). Comparative effectiveness of composition skills learning activities in the secondary schools (Cooperative Research Project 1528). Madison, WI: University of Wisconsin (ERIC Document Reproduction Service No. ED 003 279).

McCurdy, J. (1981, December 9). California professors recommend programs to improve skills of college-bound students. The Chronicle of Higher Education, 23(15), 1.

Murray, D. M. (1978, October). Teach the motivating force of revision. English Journal, 67(7), 56-60.

National Commission on Excellence in Education (1983). A nation at risk. Washington, D.C." Government Printing Office.

Odell, L. (1974, Summer). Measuring the effect of instruction in prewriting. Research in the Teaching of English, 8, 228-240.

Perl, S. (1979, December). The composing process of unskilled college writers. Research in the Teaching of English, 13(4), 317-336.

Pianko, S. (1979, February). A description of the composing process of college freshman writers. Research in the Teaching of English, 13(1), 5-22.

Scully, M. (1981, December 8). Entrance rules tightened at some public institutions. The Chronicle of Higher Education, 23, 1.

Shaughnessy, M. P. (1977). Errors and expectations, A guide for the teacher of basic writing. New York: Oxford University Press.

Stallard, C. K. (1974, Fall). An analysis of the writing behavior of good student writers. Research in the Training of English, 8, 206-218.

Wolf, M. H. (1966). Effect of writing frequency upon proficiency in a college freshman English course. (Cooperative Research Project 28446). Amhers, MA: University of Massachusetts. (Eric Document Reproduction Service No ED 003 384).

John E. Creeden is Assistant Dean of the College of Letters & Science, University of Wisconsin-Madison, Madison, Wisconsin 53706

ACADEMIC ASSESSMENT: IMPLICATIONS FOR FACULTY EVALUATION

Robert DeBard, Ed.D.

Think about this title. If the word "faculty" was to be replaced with the word "student," the subject would not only cease to be controversial, but the title would also be accused of being redundant. After all, isn't academic assessment just a 1980's bit of jargon that really means more strenuous student evaluation? There is little debate as to our right to assess the achievement of our students on any set of variables the professional members of the academy deem fitting. Whether it be the national standardized tests used by Northeast Missouri State University, the student performance on assigned tasks or in decision-making situations of the eight basic "skills that last a lifetime" used by Alverno College, or the more common course-based examinations used throughout the ages by thousands of institutions several times a semester, the right to assess student knowledge and assign evaluation symbols accordingly is a well established part of the ethos of higher education.

What I wish to explore today is more provocative in its implications for the practice of higher education. Academic assessment can have far reaching impact on the evaluation of faculty members as well as students. In my study of the current literature associated with the assessment of academic achievement movement, however, I have discovered a dearth of reference to faculty evaluation. There is constant reference to various ways of evaluating student performance such as "value-added" or "attainment" measures. It is often cited that one of the benefits of effective assessment can be curricular reform. There is far less reference, but still some, regarding faculty self-assessment based upon various findings. But I have yet to read about a program that boldly states that academic assessment is going to be a part of the reward system, including promotion and tenure, of faculty at a given college or university. Because of the potential for this to be done in an external as well as internal manner, I think it best for some consideration to be given to the implications of academic assessment on faculty evaluation.

I must confess that I come to you as a provocateur rather than an expert in either academic assessment or faculty evaluation. As Dean of Student Development at Old Dominion University, having served for four years as dean of a school responsible for freshman services, including diagnostic testing, academic skills development, and advising, my bias is with the student as a consumer. I also serve as the administrator of undergraduate continuance regulations, which forces me to confront student academic failure with some regularity, and I have come to realize that there is a certain constancy of subject matter, learning format, and assigned professor that come together all too often, resulting in low student evaluation, albeit in an objective manner. Balancing this perspective is the fact that I spent four years on the University's Academic Vice President Tenure and Promotion Review

Committee, which provided me with an interesting view of how faculty are evaluated when it really counts. Finally, my School submitted the incentive grant that was accorded to Old Dominion University for its Assessment of Academic Achievement Program. So while I am vitally interested in the assessment of academic achievement, my responsibility for faculty evaluation is of no great consequence. Still, I serve up these thoughts to provoke discussion about an aspect of the assessment of academic achievement movement that is bound to result in one form or another. My thinking is that the sooner those of us in higher education consider it, the better the chance we have of controlling its outcome.

To appreciate why the assessment of academic achievement movement could result in some impact upon faculty evaluation, let's consider the roots of the movement itself. The NIE's Involvement in Learning, the AAC's Integrity in the College Curriculum, and the SREB's Access to Quality Undergraduate Education all depict deficiencies in the undergraduate curriculum and apparent outcomes and propose assessment of academic assessment as a primary need in the reform. These are indeed reform documents, long on the negative aspects of how undergraduate education is currently being practiced. Within the state of Virginia, there has been a move to pick up on this reform. A report entitled The Measurement of Student Achievement and the Assurance of Quality in Virginia Higher Education was the result of a study that was a part of Resolution 125. While the study itself is quite constructive, the resolution implies the potential for negative interpretation:

> WHEREAS, Virginia's public institutions of higher education are a source of pride to the Commonwealth and the basis for the State's continued economic and cultural growth; and
>
> WHEREAS, Virginia has an investment in excess of $1 billion in physical plant and over $300 million in equipment in its institutions of higher education; and
>
> WHEREAS, Virginia historically devotes over seventeen percent of its general funds in the biennial budget to higher education, which amounts to over $1.3 billion in general funds in the current biennium; and
>
> WHEREAS, continued broad public support for Virginia's system of higher education is essential to the system's growth and well-being; and
>
> WHEREAS, various studies of higher education have raised questions about curriculum requirements, quality of instruction, and student achievement in the nation's colleges and universities; now, therefore, be it
>
> RESOLVED by the Senate, the House of Delegates concurring, that the State Council of Higher Education conduct a study on student achievement in Virginia's public higher education system, and to investigate means by which student achievement may be measured to assure the citizens of Virginia of the continuing high quality of higher education in the Commonwealth.

It is not my mission to discuss why all this attention is being given to the evaluation of higher education at this point. Suffice to say that it will not go away in the near future, so the implications have to be confronted. Institutions seem to be willing to do so, insofar as the evaluation involves students or even curriculum. Reform of general education has been, or is being, studied in an estimated two-thirds of all four-year institutions (Carnegie Foundation for the Advancement of Teaching). The problem has been discussed by Peter Ewell in Change magazine as well as his other considerable writing in the area:

> Classic student assessment efforts tended to be both institution and curriculum-specific. Their primary objective was to serve as an additional mechanism for gauging an individual student's mastery of particular body of knowledge in order to provide guidance for future development. In contrast, most current discussion revolves around the curriculum or institution as the unit to be evaluated. Assessment results are used to support judgments about the effectiveness of particular programs and institutions in meeting stated institutional goals...Assessment for the purposes of curriculum and program evaluation is, by definition, external to the educational process; it may be undertaken to improve teaching and learning, but it is not actually about teaching and learning. This means that the responsibility for the assessment process lies as much with academic administrators as with faculty - a notion that, quite naturally, makes many faculty uncomfortable. (Ewell, Change, November/ December, 1985)

Let's consider how faculty are currently evaluated as compared to how the assessment of academic achievement movement would seem to suggest how teaching and learning should be evaluated. During my tenure on the Promotion and Tenure Review Committee, I had access to all documents and evaluations that were turned in on various faculty from department chair, department, school, and faculty senate committees, as well as their deans. I found the overall process comprehensive, but ultimately soft. Especially when it came to the evaluation of teaching. Research, whether published in scholarly journals or books or as evidenced in external funding, was not hard to quantify, although there did seem to be trouble with quality standards. But teaching itself, offered a dearth of evidence. Some mention was made, on occasion, to student evaluations that involved "popularity" variables that gauged student contentment more than achievement. Though there was peer review of publications, there was no evidence of peer review of teaching. Indeed, the whole evaluation process seemed more to smack of "thumbs up or thumbs down" at all levels. Department chairmen, when they commented at all, usually made vague references to needing more scholarly production or expressed some appreciation for departmental committee work. In the ninety-four cases that were reviewed during my tenure, not one reference was ever made to an individual faculty member's teaching effectiveness as evidenced by student achievement.

The reason why this is a problem is that the onslaught of assessment of academic achievement is going to provide some ability that has apparently been missing heretofore. There still might be the motivation on the part of the faculty evaluators to use the data that will be accumulated, but the data will almost certainly be available for any group intent upon such evaluation.

My experience tells me that faculty will be slow to accept this potential and take hold of the practice. I chaired a self-study of remedial education a few years ago, in which the academic skills of writing, reading, and mathematics were considered. Old Dominion University has developed a fairly impressive pre-post test capability within the developmental studies area. In my report, I had the temerity to suggest that this same pre-post test concept should be instituted within the academic disciplines and that findings should be used to judge the effectiveness of teaching as well as learning that had taken place. Suffice to say, the reaction was instant and volatile. Faculty seem to reserve the right to judge students, but they are less than reluctant to be judged upon the performance of their students. This has been a particularly troublesome part of lower division education in mathematics. Take, for instance, our first course in calculus. Keep in mind that only the students intent upon an engineering, math, or computer science degree would take calculus as a natural course. Last fall, of the 349 students enrolled, the average grade given would have created a grade point average of 1.4, with 41 percent of the students withdrawing without any grade at all. One section had an average of 1.23, with only 19 of the 117 students registered receiving a "C" or above, and a dropout rate of 55 percent. Several other math courses come close to rivaling calculus for low honors. It is easy to understand the reluctance of math faculty to be judged with relation to this performance. I have no doubt as to who is primarily responsible for low performance in any given classroom. I meet up with many of these students, as I must place them on probation or suspension at the end of each grading period. But the total denial of responsibility of teaching faculty is to invite others, myself included, to form judgments that could be misinformed., Quite simply, academic achievement, or the lack of such, must be figured into any equation of faculty teaching as well as student learning.

Since I did cite a negative aspect of current practice at Old Dominion University, I would like to shift to a more positive area related to the Assessment of Academic Achievement Proposal that has been funded. We have proposed to develop tests that would assess the value-added learning that takes place in the areas of writing, mathematics, reading, and critical thinking. We have already developed diagnostic tests in the areas of writing and mathematics, as well as use of the Test of Reading Power of The College Board. We will study various tests of critical thinking that have been developed, as well as consideration of our own. What will be far more difficult will be the development of proficiency tests that could conceivably be used as exit exams. We know the difficulty because we have already achieved an impressive level in the area of writing through the administration of our Writing Sample

Placement Test to all incoming students, and the Exit Exam of Writing Proficiency to all students who may start taking the test after they achieve their junior status. One of the first things that will need to be determined is just what use will be made of these tests once developed and administered. There has not been any reference made to faculty evaluation because we did want the proposal to leave the University, but the potential is there.

What has been built into the proposal is the heavy use of faculty consultation in the development, appraisal, and implementation of the proposal. We learned this through the successful development and implementation of the Writing Sample Placement Test and the Exit Exam of Writing Proficiency. A task force of faculty from each of the academic schools was formed. One of the early decisions, however, was that the ultimate placement of the responsibility for the administration, grading, data collection, and analysis of the writing sample as well as the exit exam would not rest in any one regular school or discipline, but rather in a unit without a vested interest in FTE generation or political rivalry. It is hoped that the same will be done for the math, reading, and critical thinking areas, although there are some serious questions currently being asked because of a recent University reorganization.

Another essential element of the Assessment of Academic Achievement Program is that amount of time and resources being placed into the development of a computer program that will be able to calibrate test scores, trace pre-post test improvement, and retain data by student, subject, and curriculum taken. Those of you who have been involved in trying to develop such a system know that the labor involved is not to be underestimated. We have actually built money for a programmer into the budget rather than expect our overstrapped Computer Center to do these good deeds with existing staff. I also think you can see the potential for using these data to trace development through specific curriculum and the faculty assigned to teach this curriculum.

If this is to have any chance of working, however, it will have to be viewed in a positive light by faculty. Here again, we can use the development of the exit writing exam as a lesson in what to do. The key, again, has been faculty involvement and ultimate control. The ad hoc task force was turned into an ongoing standing committee whose membership served for limited periods of time and cut across discipline lines while always involving members of the Basic Writing and English Composition Programs.

The exit exam, itself, used essay questions that were gained from each discipline s faculty, to be applied to the major area of concentration for students. This addressed both the question of faculty input and subject matter relevance. We place emphasis upon the fact that these are the disciplines of students who are taking the exam, and the performance on it is a reflection of the literacy to be found within the majors in the discipline. While this has resulted in some negative remarks about our engineering students, it has instilled a feeling of ownership and responsibility on the part of faculty. We have been able to trace a direct correlation between the number of writing courses a given student has taken going into the exam and the percentage of pass rates. This might seem obvious, in that liberal arts majors would be expected to perform better than their technology or science counterparts, but it has had some implications for electives and general education courses.

Even more important has been the development of a Writing Across the Disciplines Program. Another finding is that there is also a direct connection between the length of time that has passed since a student has taken a course that has had essay writing as a primary evaluation tool and the percentage of failures on the initial exit exam. Since the upper division curriculum of most disciplines is rather prescriptive, especially in the professional schools, the opportunity to build writing courses themselves into the curriculum is very limited. The alternative is to build writing into course matter that previously did not incorporate such writing into student performance. Though the findings are still preliminary, indications are that there is a correlation between students who take courses involved in the Writing Across the Disciplines Program and pass rates on the exit exam.

The University has taken a very positive, incentive driven, approach to this Writing Across the Disciplines Program. The first two summers of the Program provided paid professional development for discipline leaders who were willing to study how they could incorporate writing into their subject areas. These leaders were then asked to work with other interested faculty. The Director of the Old Dominion University Writing Center is provided release time to work with individual faculty who wish to become involved. While there is nothing negative accorded to not becoming involved, there is positive feedback to department chairs on behalf of faculty who have been active.

I recount this success because it might indicate what would appear to be the key to the eventual success of tying faculty evaluation to student achievement. As long as it is viewed as a threat, it will be rejected by faculty who, quite frankly, have a cartel when it comes to peer evaluation. If, on the other hand, evidence of successful student achievement can be positively tied to faculty teaching, the involved faculty will welcome the assessment of academic achievement as an ally with regard to their own evaluation.

But I do not mean to paint a totally rosy picture here. I fear that faculty will be reluctant to join the movement, and such reluctance could cast doubt in the minds of external bodies who would gladly make judgments on behalf of the faculty. There is a constitutional rule that allows a person not to testify against himself. My lawyer friends tell me that the effectiveness of exercising this option is tied to the perception of guilt or innocence on the part of the jury before any such testimony would take place. The greater the perception of guilt, the greater the need for testimony. As Bob McCabe of Miami-Dade indicated as a part of a panel discussion the the American Association of Higher

Education meeting in Washington last March, "Make no mistake, the assessment movement should be viewed as a criticism to find out how bad rather than how good we are doing." I would concur, insofar as external bodies control the system. Given what the state of Florida has instituted, I think we can appreciate how the President of Miami-Dade Junior College might feel. But for most of us, there is still time to take the initiative and assume our proper responsibility. I credit the state of Virginia with a healthy approach thus far. The preamble to the recommendation of our State Council states it better that I could:

> Citizens of the Commonwealth, parents, and students have a right to hold institutions of higher education accountable for effective teaching and learning. Institutions can benefit from more systematic knowledge of student achievement. The Council recommends against a system-wide minimum competency testing program for Virginia as the best means to measure student achievement. This approach lacks several characteristics of a good assessment program. It promotes standardization, is insensitive to important institutional differences in mission and curriculum, ignores broader educational objectives which should be assessed, threatens to establish minimums as the norm, and fails to contribute to improvement of the teaching and learning that occur in classrooms. Virginia has worked long and carefully to nurture a diverse set of colleges and universities. It is not appropriate to impose a single statewide test upon this diversity. (November 20, 1985)

Within this positive context, I advance the premise that the assessment of academic achievement could lend as much credibility to excellent teaching as it could result in criticism. I might be naive in assuming that what takes place in the average classroom is underrated and evaluation is based upon an absence of good data. There is currently the energy and resources being made available to find out what is happening in the classroom, and it behooves teaching faculty to be willing to have good teaching documented and thus accepted as an integral part of good learning.

Robert DeBard, Ed.D. is Dean of Student Development at Old Dominion University, 1411 West 49th Street, Norfolk, Virginia 23508

ASSESSMENT OF GENERAL EDUCATION OUTCOMES
CLAYTON STATE COLLEGE

James Doig, Linda Greer and Joyce Swofford

The development of a coherent general education program which is focused on student outcomes and provides for systematic assessment of student progress toward those outcomes has been the major initiative of the Clayton State College faculty during the past three years. The goals of this work may be summarized as follows:

1. To identify and clearly define the general education outcomes that should occur for all graduates of Clayton State College--i.e., the common abilities and knowledge for all students;

2. To determine and implement the curriculum structure, course content, and instructional techniques most likely to enable students to be able to demonstrate these outcomes; and

3. To develop and implement a comprehensive system for assessing student performance at the beginning, during, and at the end of the student's program.

The basic features of the program are summarized below.

Student Outcomes

General education will be defined in terms of student outcomes, i.e., the common skills and knowledge that all students should acquire as a result of general education. These outcomes will be expected of anyone enrolled in a baccalaureate or pre-professional degree program (those programs which require 60 quarter credit hours of general education courses). To the extent appropriate, the outcomes will also be expected of students in career and technical programs.

The faculty has identified eight outcomes that are important for all students to acquire and to be able to apply and integrate as a result of their general education experience. The outcomes include two skills and six perspectives:

Skills

Communication: The ability to express and understand meaning through writing, speaking, reading, and listening.

Critical thinking: Information and knowledge gathering and reasoning characterized by careful, continual thought.

Perspectives (knowledge-based frames of reference within which one understands, interprets, and evaluates what he encounters)

Aesthetic Perspective: The perception and interpretation of the aesthetic potential inherent in the interaction between an individual and an object, work, or experience.

Contemporary Perspective: The application of knowledge of institutions, ideas, values, forces, trends, personalities and events to the interpretation and evaluation of the various spheres of the contemporary world--economic, political, religious, social, cultural, environmental, scientific and technological.

Historical Perspective: The ability to view facts and develop rational expectations from an understanding of the past.

Mathematical Perspective: The ability to use knowledge of mathematics to understand, manipulate, interpret, evaluate, communicate, and apply information that can be expressed in numerical or other related abstract symbol systems.

Scientific Perspective: The ability to examine phenomena and to interpret everyday experiences in an orderly and logical manner consistent with the principles and theories of science.

Value Perspective: The ability to identify values and their sources, analyze a value conflict situation, and make value-based decisions.

This set of outcomes is not unique. The goals and concepts implicit in the Clayton State College general education outcomes are similar to those of many institutions and are recommended in many studies of general education. The careful, detailed descriptions which have been developed for each of the outcomes, however, do not seem characteristic of other statements.

Each of the outcomes is described in terms of its components and the ways (modes) in which it is expressed. For example, the components of critical thinking have been defined as information/knowledge gathering (observing and clarifying) and reasoning (organizing, analyzing, synthesizing, and evaluating). Modes in which these component skills are expressed include problem-solving/decision making and inquiry. The outcome descriptions are further extended and clarified for each mode by the criteria that have been developed for assessing student performance. These descriptions represent institutional definitions that promote consistency and common understanding among students and faculty of desired program outcomes. They also provide the basis for general education instruction and assessment.

Outcome-Focused curriculum

The faculty is developing a new outcome-focused curriculum that will replace the existing general education curriculum. "Outcome-focused" means that the purpose of the curriculum is to help students develop the desired skills and perspectives. In selecting an appropriate set of experiences, the primary considerations must be:

- providing students the opportunity to acquire and apply all eight outcomes;

- providing the essential concepts and knowledge that are required for the acquisition of the skills and perspectives and for their application outside of general education courses; and

- encouraging and requiring integration of the outcomes.

What will be the place of the disciplines in an outcome-focused curriculum? The unique knowledge and approaches which characterize the various disciplines will provide the breadth and depth of content and experiences necessary to help students develop the skills and perspectives. In designing the curriculum the key questions are, "How can the disciplines contribute to enabling students to develop the skills and perspectives?" and "Which discipline content is most important and will be most effective in doing this?" Because the outcomes are not associated solely with a particular discipline or course but are developed across the curriculum, the outcome-focuses approach enhances the coherence of the general education program as it is taught and as students experience it.

The College's current general education curriculum for transfer program students, like other institutions in the University System of Georgia, consists of twenty quarter hours of studies in each of three areas: the Humanitities, the Natural Sciences and Mathematics, and the Social Sciences. The basis for the content of courses required of all students as well as the array of courses from which students may choose in fulfilling the distribution requirements is primarily that of exposing students to particular disciplines. The outcome-focused approach to general education requires the faculty to completely rethink this curriculum. The potential impact of the outcome-focused approach is suggested by the new program for writing and speaking instruction implemented this past Fall and described in Section IV of this report.

The first phase of the process of translating the set of eight outcomes into a general education curriculum--the development of initial proposals for courses and other educational experiences--has just been completed. The plan is to implement a new curriculum in the Fall of 1987.

Assessment

In order to provide frequent feedback and to document student progress in developing the eight outcomes, assessment is to be an important and integral part of the general education program. Assessment will occur throughout a student's program of study.

Entry Assessment. Students will be introduced to the assessment process for some outcomes before registering for their first courses at the College. For other outcomes, entry assessment will take place after intitial enrollment. The purpose of entry assessment is two-fold:

(a) For some outcomes (e.g., Communication) assessment results will determine placement into appropriate educational experiences.

(b) For other outcomes, entry assessment is important in making the student and faculty aware of the student's starting point. This information will assist students, faculty advisors, and instructors in determining the educational experiences that will be of most benefit and will also provide indicators of progress students make during their time at Clayton.

Assessments during the Program of Study. Students will be provided with a variety of opportunities to demonstrate performance relative to each skill and perspective.

(a) Successful performance on assessments in required courses or other experiences which provide the basic instruction related to a specific skill or perspective will be a requirement for successfully completing these courses (e.g., demonstration of the required level of writing and speaking is required to receive a passing grade in Freshman English courses.)

(b) Since the ability to use the skills and perspectives in a variety of settings and tasks is important, assessments of all outcomes will occur in different courses and disciplines across the curriculum. Because the skills and perspectives are typically used in combinations of two or more, most assessments will involve the integration of outcomes. Most of these assessments will take place within a course, but some will be external to specific courses.

Graduation Assessment. As students approach graduation, they will need to demonstrate that they can integrate and use the skills of communication and critical thinking and a number of the perspectives to accomplish a significant task. In many cases, this task will be related to the student's major area of study.

Multiple assessments of the same outcome will be critical not only in reinforcing student learning but also in enhancing the validity of assessment information. The use of common criteria and standards for effective performace in assessing an outcome will reinforce the student's understanding of the outcomes as well as his or her sense of the coherence of the general education experience.

This approach to assessment (including placement based on entry assessment, required performance standards for successfully completing the Freshman English course sequence, and the use of common criteria for assessing writing across the curriculum) has been imlemented for writing. Partial implementation for speaking has also taken place. Assessment criteria have been developed for each outcome. The criteria provide an extension of the outcome descriptions as well as the basis for the development of an assessment process based on the approach described above. In addition, example assessments for some outcomes have been developed, primarily as a means of further clarifying the outcome.

THE COMMUNICATION PROGRAM

The College implemented an outcome-focused, assessment-based program in writing for students entering the College in the Fall Quarter, 1986. Initial steps were also taken to incorporate speaking into the program. The features of this program and the processes through which it was developed are summarized below.

The approach to communication represented by the current program is founded primarily on the principles of traditional rhetoric. That is, it emphasizes the recognition of audience, the establishment of purpose for the communication with regard for the needs of the audience, and the necessity for flexibility in the planning and organization of a communication (rather than the memorization of formulas). It requires the ability to use standard edited English conventions because these are generally required by the variety of communciation tasks a college-educated communicator would be likely to encounter.

The portion of the program designed to provide instruction in speaking is similarly founded. It assumes that college-educated communicators are called upon to speak as often as, or more often than, they are called upon to write. Since speaking is vitally important to college-educated communicators, and since it demands that the speaker be aware of and able to use some of the same rhetorical principles as the writer (albeit in different ways and with different emphases), instruction that draws connections between the two seems to be reasonable. It is believed that combining instruction and assessment of both modes in the same course may very well enhance the student's understanding of and ability to master each.

An important feature of the program is the assessment of student writing and speaking in terms of the criteria developed by the Communication Outcome Council. These criteria with clearly described performance ratings make it possible to develop standards which are consistently applied in all sections, to communicate these expectations to students, and to teach students to apply the critera in evaluating their own writing and speaking and that of others. With clear criteria and expectations, communication between the student and teacher is enhanced, and both student and teacher are better able to judge when and to what degree the student has achieved the outcome.

Introductory Communication Instruction

The college-level "focused-instruction" in writing and speaking occurs in two five-quarter-credit-hour courses: English 111 and English 112. Both courses were developed specifically to provid the initial college-level instruction in writing and speaking as defined in the Communication Outcome description. In both courses student writing and speaking are assessed using the criteria developed by the Communication Outcome Council. The syllabi for both courses illustrate how the courses address the modes of writing and speaking. Students are strongly advised to take both English 111 and English 112 as early as possible in their programs.

For entrance into English 111, all students are required to complete a diagnostic entry writing assessment. Their ratings on this assessment, along with other information which is part of the placement process mandated by the University System of Georgia, is used to determine a student's placement recommmendation for English courses. Some students are advised into remedial courses in English on the basis of this assessment, while others are advised into a supplemental writing laboratory taken concurrently with English 111 or English 112. Although no pre-college entry assessment in speaking is administered, students in English 111 and English 112 are assessed in speaking during the first few days of the courses. On the basis of these assessments, students may be referred to a supplemental speaking laboratory.

Students' exit from English 111 and from English 112 is based on their passing the writing exit assessment. Students may, under some circumstances, exit the course with a minor deficiency in the skills measured by the writing criteria, provided they continue in the writing laboratory as appropriate until they meet or exceed the required standard. At present the exit assessment in speaking is incorporated into the grades for the courses.

Writing-Across-the-Curriculum

As a result of the recent attention to the Communication Outcome, the emphasis that most faculty members at the College have traditionally placed on writing seems to have increased. A major difference in approach is the adoption of a campus-wide definition of effective writing (i.e., the outcome description, criteria, and performance standards).

The outcome-focused, assessment-based approach provides for assessment of writing, based on common criteria and performance standards, across disciplines. Although such assessment should take place whenever students write, it must also occur as a formal requirement in order to document and certify student performance. In keeping with this approach, the Communication Outcome Council's current proposal for across-the-curriculum writing assessment would require students to formally demonstrate writing proficiency in four settings after completing the English 111-112 sequence. Students who have difficulty meeting this requirement will enroll in an individualized writing laboratory to receive assistance. If the proposal is accepted by the General Education Council and then by the Educational Policies and Executive Committees, the policy will be in effect for students who entered the College during the Fall Quarter, 1985, or subsequently. Meanwhile, because many faculty are using the writing criteria, informal across-the-curriculum assessment is already occurring.

Jame Doig is Professor of Philosophy, Clayton State College.

Linda Greer is Director, Institutional Research, Clayton State College.

Joyce Swofford is Assistant Professor of English, Clayton State College.

PERFORMANCE FUNDING FOR TENNESSEE HIGHER EDUCATION: HOW DOES IT WORK AND WHERE DID IT COME FROM?

Richard G. Dumont

Introduction

In February of 1984, the Tennessee Legislature adopted the Comprehensive Educational Reform Act, which was intended primarily to affect dramatic improvements in public schools, but which also included explicit and far-reaching provisions for public higher education. The Act provided ten million dollars for the initiation of the "Centers of Excellence Program," and it contained a list of several five-year improvement goals. Among the goals set for the public universities were increases in the following:

ACT/SAT scores of entering students;
Percentage of entrants earning degrees;
Standardized examination scores of graduating seniors;
Scores on the National Teachers Examination;
Passing rates on all parts of professional licensing examinations on the first attempt;
GRE scores of graduate schools entrants; and
Measured knowledge of graduate/professional school completers.

The Act further charged the Tennessee Higher Education Commission (THEC), the statewide coordinating agency, with the responsibility for developing and monitoring quantitative indicators or measures of achievement for each of the goals. The provisions of the Act and the monitoring role assigned to the THEC were natural outgrowths of the State of Tennessee's prior ten year history of experimentation with and implementation of performance funding for higher education in the State.

Performance funding in Tennessee and the passage of the Comprehensive Educational Reform Act and subsequent implementation of its provisions are worthy of careful attention and monitoring by both educators and public officials because of their unique potential for informing future decisions in Tennessee and elsewhere regarding the following pervasive concerns and issues:

1st. The nature of the relationship between the State and its higher education institutions with regard to shared authority for assuring accountability, and the closely related issues of institutional autonomy and academic freedom;

2nd. The role of the statewide coordinating agency as intermediary or bridge between the State and the institutions;

3rd. The increasing attractiveness of utilizing "quality indicators" or "outcomes measures" to demonstrate accountability;

and 4th. The exploration of the desirability and feasibility of basing at least some portion of State appropriations on a performance effectiveness criterion.

Recent actions and events have placed the State of Tennessee in a rather unique position nationally. While there is evidence of increasing interest and activity in several states (such as New Jersey, Florida, Georgia, Maryland, Arkansas, Colorado, and others), no other state has established such broad and prescriptive legislative authority over planned or expected instructional outcomes of public higher education institutions; nowhere has the role of a statewide coordinating agency been so precisely defined to include the development of specific quality indicators based upon outcomes; no state has yet embarked upon such an ambitious program of outcomes assessments; and none has so clearly linked a significant portion of the state appropriation to performance on outcomes measures. Although it is certainly true that most states have evidenced increasing interest in assuring accountability in public higher education over the past fifteen to twenty years, the tendency has been to focus almost exclusively on the traditional indicators of quality, "focusing on process variables (e.g., faculty, institutional resources) and input variables (e.g., student characteristics)" (Green, 1981, p. 74). As John Folger has indicated, states have characteristically been concerned with "procedural" rather than "substantive" accountability (1984, pp. 4-5).[1]

To what degree instructional and institutional effectiveness will show demonstrable long-term improvement as a result of performance funding and "Bragg Marks" in Tennessee yet remains to be seen. Similarly, it is too soon to predict whether Tennessee might reasonably be expected to serve as a model to be emulated by other states. Whatever the effects within and outside of the State, however, the case of Tennessee warrants careful examination and study because it may be seen as a kind of "natural experiment," (Babbie, 1983), a testing ground for contemporary and emerging accountability notions and procedures.

Accordingly, our purpose here is not to argue for or against the "attractiveness," "desirability," or "appropriateness" of one state's attempts to improve the quality of education for its citizens; but, rather simply to seek an understanding of how and why Tennessee has come to find itself in its unique position. Such an understanding is perhaps best achieved by an historical review of the significant events which preceded and directly influenced the evolutionary unfolding of performance funding and legislated instructional and institutional outcomes for Tennessee higher education.

In describing background factors, the author draws upon his experiences as a participant in several activities and capacities during the period February, 1977 through August, 1984. While on the faculty at Tennessee Technological University, he served as Director of that institution's Performance Funding Pilot Project, a role which also brought him into close familiarity with the statewide performance funding project effort (Dumont, 1978; 1980). Subsequently, he also served as Director of the State Board of Regents' Student Assessment and Program Evaluation Program, and the State Board of Regents Universities' representative to the THEC's Instructional Evaluation Variables Advisory Committee, the committee which was responsible for reviewing and revising the original performance evaluation component of the funding formula. Although the principal concern is with the description of events occurring at the statewide level, selected experiences at Tennessee Tech are included for purposes of illustration.

The Performance Funding Project

The Performance Funding Project was a four-year statewide effort initiated in 1974. It was sponsored by the THEC and funded by grants from the Fund for the Improvement of Postsecondary Education (FIPSE), the W.K. Kellogg Foundation, the Ford Foundation, and a donor wishing to remain anonymous. The primary purpose of the project was to determine the feasibility of performance funding as a complement to a more conventional enrollment-based formula. The intent of the project was articulated explicitly as follows: "To explore the question of whether it might be possible to allocate some portion of state funds to community colleges and universities on a performance effectiveness criterion (how good) as compared to the current credit hour and enrollment criteria (how much)." (Bogue, Harris and Troutt, 1976, pp. 1-2).

Additional clarification of the desired outcomes of the Performanoe Funding Project is made possible by a consideration of certain "boundary conditions," which any suggested revision of the existing funding policy would need to satisfy. Any such changes to the funding formula must.. "(1) be politically acceptable, easily understood and accepted by legislators and members of state government; (2) be professionally acceptable, striking the right balance between the need for institutional autonomy and the need for state-level review; (3) encourage institutions to exercise initiative in developing performance measures on which they might eventually be funded; and (4) recognize differences in institutional role and environment and promote diversity." (Bogue and Troutt, 1977, p. 4).

The execution of the Performance Funding Project in Tennessee involved several significant activities and accomplishments (see EXHIBIT 1), including the establishment and work of a National Advisory Panel and a Statewide Advisory Committee of participating governing boards, institutions, legislators and officials., visits by THEC project Staff to public institutions in the State; the use of external consultants; a survey of other states' activities; visits to centers of other related activities (e.g., NCHEMS, ACT and ETS); and the successful pursuit of funding support; to cite some of the more significant (Bogue, Harris and Troutt, 1976, pp. 3-4).

Of the several activities, however, it is likely that the most significant and most critical to project success were the pilot projects conducted at eleven selected institutions from throughout the State

EXHIBIT 1*
THE PERFORMANCE FUNDING PROJECT IN THE STATE OF TENNESSEE
1974-1978

SPONSORSHIP AND FUNDING:
- The Tennessee Higher Education Commission (THEC)
  - The Ford Foundation
  - The W. K. Kellogg Foundation
  - FIPSE
  - A Donor wishing to remain anonymous

PURPOSE:
- "To explore the questions of whether it might be possible to allocate some portion of state funds to community colleges and universities on a performance effectiveness criterion (how good) as compared to ... credit hour and enrollment criteria (how much)."

PROCEDURES AND PARTICIPANTS:
- Phase I - Planning - 1974 - 1976
  - National Advisory Panel
  - State Advisory Committee
  - Appointment if Project Staff
  - Securing Foundation Support

- Phase II -- Campus Based Pilot Projects - 1976 - 1977
  - State Board of Regents System
  - AustinPeay State University
  - Columbia State Community College
  - Memphis State University
  - Shelby State Community College
  - Tennessee Technological University
  - Volunteer State Community College

- University of Tennessee System
  - UT Center for the Health Sciences
  - UT at Chattangooga
  - UT at Knoxville
  - UT at Martin
  - UT at Nashville

- Phase III - Pilot Project Completion/Funding Policy Options - 1977 - 1978

- Phase IV - Funding Policy Implementation/Project Dissemination - 1978 - 1979

*Sources: Bogue and Brown (1982); Bogue, Harris and Troutt (1976); Bogue and Troutt (1977) (1979); Dumont (1978) (1980).

(Bogue and Troutt, 1977). The major objectives of each of the eleven pilot projects were as follows:

"First Year - 1976-1977. The identification and/or development of institution-wide instructional goals (general education goals) and associated performance indicators."

"Second Year - 1977-1978. The acquisition of data on the indicators and the development of ways by which performance might be incorporated into the State funding formula." (THEC, 1976).

The Pilot Project at Tennessee Technological University (see EXHIBIT 2)

A representative group of university faculty members were appointed to a project steering committee as was a director (the author) for the project. Following a comprehensive review of available literature pertaining to general education goals, and after much discussion and debate, the steering committee developed a list of over thirty possible goals, which were subsequently included in a campus-wide goals survey. The results of the survey yielded a finalized set of twelve major general education goals for the institution, clustered into the areas of communication skills, basic understandings of science and technology, basic understandings of history and the social sciences, problem solving abilities, and preparation for further study and/or employment.

The project steering committee also identified performance indicators for the several goals, including among the more important the ACT COMP (College Outcomes Measures Project) and student and alumni satisfaction surveys, based upon NCHEMS models. During the months of January through April of 1978, the outcomes assessments were conducted, utilizing representatives samples of graduating seniors and alumni (Dumont, 1978; 1980).[2]

In accomplishing the project objectives, significant resistance was encountered and concerns were expressed by members of the university community, particularly faculty. In the initial planning stages and throughout the two years of the project, there were oft-repeated concerns regarding: (a) the potential for increasing control and internal "meddling" by the THEC and the governing board; (b) the likelihood of the development of pressures for the indicators, which are the "means" of evaluation, to become "ends" in themselves, thereby displacing the corresponding instructional goals; (c) the "uninformed" interpretation and use of performance data; and (d) the apparently, politically and practically insoluble "rich get richer" and "poor get poorer" vs. "Robin Hood" dilemma of performance-based funding.

With few exceptions, the trepidations and objections raised and articulated at Tennessee Tech University were in evidence at other pilot project institutions. That the project was successful, despite clear and legitimate misgivings and strong and vocal resistance was due to various agreements and accommodations reached between the institutions and their governing boards and the THEC (for details, see Dumont, 1980, especially pp. 411-414).

The Performance Funding Policy for Tennessee Higher Education (see EXHIBIT 3)

The success of the pilot project at Tennessee Tech University, as well as at several of the other participating institutions, accounted in considerable measure for the adoption and implementation in 1979 of a Performance Funding Policy for the State of Tennessee. That policy allowed an institution to earn an additional amount of its E&G appropriation--up to 2%--based upon its engaging in and reporting the results of specific evaluation/assessment activities. The determination of the "performance" funded amount was based upon the number of points an institution scored on a series of five evaluation "variables:" I - Accredited Programs; II - General Education., III - Major Field; IV - Satisfaction Surveys; and V - peer Review (see EXHIBIT 3 for definitions). An institution could earn up to a maximum of 20 points on each variable for a possible total of 100 points. For example, Tennessee Technological University earned 67 points in 1979 (the highest number of points earned that year). Since its state appropriation at the time was $12.5 million, it was eligible to receive up to an additional 2% of that amount, or $250,000 (.02 X $12.5 million). Tennessee Tech's accumulated 87 points resulted in an actual performance-based augmentation of $167,000 (.67 X $250,000).

Since its initial formulation, the State Performance Funding Policy has undergone significant modification and revision (see EXHIBIT 4). The policy has been revised and refined on the basis of experience with and perceived inadequacies of the original variables. The major changes have been in a more precise definition and specification of the variables and standards for performance, the introduction of alternative options within variables so as to accommodate institutional diversity, and the differential weighting among variables (whereas initially all variables were worth 20 points, the current weights range from a low of 5 to a high of 30). An additional major revision has been the deletion of the old Variable V, "Peer Review," and the adoption of the new "Planning for Instructional Program Improvement" variable. Also, the performance funding percentage of the funding formula has been raised to 5%, hardly an insignificant amount for most institutions in the State. Finally, and very importantly, beginning in 1984 funds were awarded based upon actual performance on variables, rather than simply upon engaging in evaluation/assessment activities.

Discussion

Whether the example set by Tennessee with regard to performance funding and the legislation of prescribed instructional outcomes will be followed by other states remains to be seen. As John Folger has stated, "The approach taken by Tennessee is likely to be appealing to a number of legislators;" but, he goes on to say, "They legislators are likely to propose goals which appear reasonable to legislators, but which pose

EXHIBIT 2*

THE CAMPUS BASED PILOT PROJECTS

OBJECTIVES:

- First Year - 1976-1977
    - The identification and/or development of institution-wide instructional goals (general education goals) and associated performance indicators.

Second Year - 1977-1978

- The acquisition of data on the indicators and the development of ways by which performance might be incorporated into the state funding formula.

THE PILOT PROJECT AT TENNESSEE TECHNOLOGICAL UNIVERSITY

- Selection of Project Director and Steering Committee
- Development of General Education Goals by the Steering Committee
- Campus-Wide Survey of General Education Goals
    - Communication Skills
        - Writing
        - Reading
        - Speaking
        - Mathematical
    - Basic Understandings of History and of the Social Sciences
        - Democracy and Citizenship
        - History and Geography
        - Economics
        - Behavioral Sciences
    - Basic Understandings of Science and Technology
    - Problem-Solving Abilities
        - Reasoning and Critical Thinking
        - Informed Acquaintance with Major Methods of Inquiry
        - Preparation for Further Study and/or for Employment
    - Identification of Performance Indicators
        - External-Validated Standardized Tests
            - ACT Assessment - Composite and Subtest Gain Scores
            - ACT COMP (College Outcome Measures Project)
        - Student and Alumni Surveys
        - Selected Institutional Activity Data

*Sources: Bogue and Troutt (1977); Dumont (1978) (1980); THEC (1976).

EXHIBIT 3*
THE PERFORMANCE FUNDING POLICY
ADOPTED BY THE THEC IN AUGUST, 1979

PRIMARY FEATURES:

- An institution may earn an additional amount - up to 2% of its budget - based on its performance (Thus, an institution having a state appropriation of $20 million could augment that amount by up to $400,000).
- The determination of the performance-funding amount is based upon the total number of points an institution scores on a series of five variables.

Variable		Maximum Points
I.	Number of academic programs accredited, such as law, engineering, education, and business	20
II.	Performance of graduates on a measure of outcomes in general education, such as ability to communicate, analyze, and evaluate, and familiarity with major modes of intellectual inquiry.	20
III.	Performance of graduates in tests in their major fields (nursing exams, engineering exams, and so forth).	20
IV.	Evaluation of programs and services by enrolled students, recent alumni, and community-employer representatives, principally through follow-up questionnaires.	20
V.	Peer evaluation of institutional programs. This entails evaluation by colleagues from similar institutions.	20
	Total maximum possible points	100

- Thus, an insitution accumulating 75 points and having a state appropriation of $20 million would receive a budget augmentation of $300,000 (.75 X $400,000) or (.75 X .02 X $20 million).
- The performance funding policy has undergone significant revision since its initial formulation

*Sources: Bogue and Brown (1982); THEC (1983).

EXHIBIT 4*
THE PERFORMANCE FUNDING POLICY REVISED
INSTRUCTIONAL EVALUATION VARIABLES AS OF JUNE, 1983

	Variable	Maximum Points
I.	Program Accredition - The number of points awarded to an institution is the percentage of the 25 points calculated as the percentage of eligible programs accredited.	25
II.	Program Field Evaluation - The institution may earn 10 points for assessing the performance of its graduates on an appropriate and approved test (externally validated or locally developed). An additional 20 points may be earned if the institution can demonstrate that the performance of its graduates show gains over time, or that they exceed the performance of graduates of similar programs at comparable institutions.	30
III.	Institution-Wide Education Outcome - Option A - Applicable to all four year institutions and may apply to community colleges. General Education Outcomes - The institution may earn 5 points for assessing a representative sample of its graduates with the ACT COMP. An additional 20 points may be earned if the institution can demonstrate that the performance of its graduates show gains in value-added over time, or that their value-added is above average as compared to the value-added of graduates of comparable institutions. - Option B - Applicable to all technical institutes and may apply to certain community college programs. Placement of Graduates - The institution may earn 5 points for conducting follow-up surveys to ascertain the employment status of its graduates. An additional 20 points may be earned by demonstrating a favorable employment rate.	25
IV.	Instructional Improvement Based on Referent Group Survey - The insitution may earn 5 points for surveying at least one of three referent groups (students, alumni or community members/employers) and utilizing the results for instructional improvement. An additional 5 points may be earned by conducting two surveys, either of the same or different referent groups.	10

EXHIBIT 4
(CONTINUED)

Variable	Maximum Points
V. Planning for Instructional Program Improvement - The institution may earn 5 points by submitting an acceptable annual plan for instructional improvement. An additional 5 points may be earned by submitting an evaluation of the degree of success experienced in executing the previous year's evaluation plan, and by demonstrating that at least half of the plan's objectives were accomplished.	10
Total maximum possible points	100

*Sources: THEC (1983). Note: The actual, current descriptions of the variables and performance standards are substantially more precise and detailed than the summaries provided here.

complex measurement problems." (Folger, 1984, p. 11). Folger further identifies a series of key problems in attempting to utilize outcomes or performance measures for purposes of accountability:

"1. Responsibility for academic quality is decentralized and diffuse;
2. Standards are determined by individual faculty and departments; they vary substantially within institutions, and even more between institutions;
3. Institutions are accountable to multiple groups including government (state and federal), students, and funders; 4. Accountability procedures at the state level are primarily fiscal and procedural rather than substantive; 5. Academic processes are diverse, and serve multiple goals; 6. There is little connection between procedures and outcomes; procedural accountability is not likely to have much impact upon the substance of education;
7. Accreditation is primarily a formative, private evaluation process. It has very little potential as an accountability process in most states; 8. Academic goals are multiple and general. Assessment of goal progress is difficult and poorly defined; 9. Quality is defined in different ways by the public, the faoulty, and college administrators; and 10. Quality assessment has important public relations consequences [in that] institutions will try to protect themselves from an unfavorable assessment." (Folger, 1984, pp. 8-9).

The above problems are clearly real and they are formidable, especially if one considers the additional complications involved in attempting to link state appropriations to institutional performance on outcomes measures.

Notes

1. Among recent contributions to the literature on the role of the state in program review and evaluation are the works of Barak (1982), Barak and Berdahl (1978), Bogue and Brown (1982), Craven (1980), Folger (1977; 1984), Green (1981), Jones (1982), Marcus (1983), and Mingle (1978).

2. For a comprehensive description of the COMP, see American College Testing Program (1980).

References

American College Testing Program. *COMP College Outcomes Measures Project: Summary Report of Research and Development 1976-1980.* Iowa City: The American College Testing Program, 1980.

Babbie, E. *The Practice of Social Research*. Belmont, California: Wadsworth Publishing Company, 1983.

Barak, R.J. *Program Review in Higher Education: Within and Without*. Boulder, Colorado: National Center for Higher Education Management Systems, 1982.

Barak, R.J. and Berdahl, R.O. *State-Level Academic Review in Higher Education*. Report No. 107. Prepared for the Inservice Education Program, Education Commission of the States. February, 1978. ED 158638.

Bogue, E.G. and Brown, W. "Performance Incentives for State Colleges." *Harvard Business Review*. November/December, 1982, 123-128.

Bogue, E.G., Harris, J. and Troutt, W.E. *Allocation of State Funds on a Performance Criterion (Performance Funding Project): A Status Report and Plan of Action*. Nashville: Tennessee Higher Education Commission, 1976.

Bogue, E.G. and Trouti, W.E. *Allocation of State Funds on a Performance Criterion, The Performance Funding Project: A Status Report*. Nashville: Tennessee Higher Education Commission, 1977.

Bogue, E.G. and Troutt, W.E. *Allocation of State Funds on a Performance Criterion, The Performance Funding Project: A Final Report*. Nashville: Tennessee Higher Education Commission, 1979.

Craven, E.C., Ed. *Academic Program Evaluation: New Directions for Institutional Research*. San Francisco: Jossey Bass, 1977.

Dumont, R.G. *Final Report for the THEC Performance Funding Project at Tennessee Technological University*. Cookeville, Tennessee: Tennessee Technological University, 1978.

Dumont, R.G. "Performance Funding and Power Relations in Higher Education," *The Journal of Higher Education*. Vol. 51, No. 4, 400-423.

Folger, J.K. *Increasing the Public Accountability of Higher Education: New Directions for Institutional Research*. San Francisco: Jossey-Bass, 1977.

Folger, J.K. "Will Colleges Have to Account to the State for Quality?" Draft of a paper presented at the Annual Forum of the Association for Institutional Research. Fort Worth, Texas, May, 1984.

Green, K.C. "Program Review and the State Responsibility for Higher Education." *The Journal of Higher Education*. Vol. 52. No. 1, 67-80.

Jones, L.V. Lindzey, G. and Coggeshall, P.E., Eds. *An Assessment of Research Doctorate Programs in the United States*. Washington, D.C.: National Academy Press, 1982.

Marchese, T.J. "Let's Reward Quality: Tennessee's Bold Experiment." *Change*. Vol. 17. No. 6, 1985.

Marcus, L.R., Leone, A.O. and Goldberg, E.D. *The Path to Excellence: Quality Assurance in Higher Education*. ASHE-ERIC Higher Education Report, No. 1. Washington, D.C.: Association for the Study of Higher Education, 1983.

Mingle, J.R. "Influencing Academic Outcomes: The Power and Impact of Statewide Reviews." In Fincher, C., Furniss, W.T., Mingle, J.R. and Spence, D.S. *The Closing System of Academic Employment*. Atlanta: Southern Regional Education Board, 1978. ED 160032.

Senate Bill No. 1. *Comprehensive Educational Reform Act*. Nashville: Tennessee Legislature. Adopted February, 1984.

Tennessee Higher Education Commission. *Performance Funding Project 1976-1978 (A Project Brochure)*. Nashville: Tennessee Higher Education Commission, 1976.

Tennessee Higher Education Commission. *Instructional Evaluation Variables and Standards 1987-88*. Nashville: Tennessee Higher Education Commission, 1986.

Richard G. Dumont is Vice President for Academic Affairs at Northwest Missouri State University, Maryville, Missouri 64468-6001

STIMULATING LOCAL ASSESSMENT INITIATIVES BY STATE POLICY: MIAMI UNIVERSITY'S RESPONSE TO OHIO PROGRAM EXCELLENCE

Curtis W. Ellison

Opportunity to earn supplementary funding through state competition offers incentive for faculty to examine student achievement -- if educational outcome information is a principal criteria for success. Encouraging local programs to establish claims for effectiveness can also stimulate self-examination, a precondition for authentic program improvement. In these respects state initiatives like Ohio Program Excellence may prove attractive to public policy makers.

Ohio Program Excellence

In appropriating $3,060,000 "to recognize and reward high qualiy academic programs at state-assisted colleges and universities" through Program Excellence awards, the Ohio General Assembly with the support of Governor Richard Celeste continued during the 1985-87 biennium an effort begun in 1983.

Program Excellence is one of five Selective Excellence initiatives addressed to a range of educational concerns. It makes one-time enrichment awards to undergraduate programs. Other initiatives are focused to stimulate research, develop branch campus and urban technical center programming, appoint eminent scholars and build base budgets of programs identified by local institutional plans. The broadest purpose of Selective Excellence is to shift attention from the expansion of higher education to increasing the quality of educational performance.

Program Excellence awards are made only to existing two- and four-year academic degree programs. As enrichment awards not to replace existing resources, funds are not a basis for on-going financial commitments. For the current biennium, awards range from $50,000 to $200,000. The number of proposals that may be submitted by an individual institution is limited by the number of degree programs offered; exemplifying their relative scale, Ohio State University in Columbus was allowed eight proposals, Rio Grande Community College four, and Miami University in Oxford six.

At the state level, proposals submitted by all institutions are reviewed in two phases by external reviewers. Tier one reviewers select proposals for scrutiny by a second panel. Tier two screening includes site visits by external consultants selected nationally to review each program and members of the Chancellor's staff. A further reduction in number results in recommendations to the Chancellor who in turn recommends to the Board of Regents; the Board makes final determination of awards. For the initial competition in 1983-84, thirty institutions submitted 131 proposals. Funds were distributed in keeping with a scored ranking, with 22 programs receiving awards up to $169,000 each.

Selection Criteria

Criteria for defining excellence include usual measures of quality such as faculty performance, data suggesting student quality, insitutional resources and support. Additional and major emphasis is placed on "evidence regarding the effects of the educational process on the students." This commiment clearly links Program Excellence to recent assessment programs nationwide, and broadens the definition of institutional quality in the direction of approaches articulated by Alexander Astin and others.

Stated criteria are grouped in three categories.

I. Program Characteristics
  A. Articulation of purpose and objectives
  B. Curriculum adequacy and appropriateness to purpose
II. Resource Characteristics
  A. Student input characteristics (SAT scores, high school GPA, number of students, etc.
  B. Quality of faculty
  C. Quality of instructional resources and supporting services
III. Program Evaluation
  A. Performance of graduates on tests in major fields where such standard tests exist
  B. Assessment of program graduates based on external measures (placement, GRE scores, national performance awards, etc.)
  C. Follow-up evaluations of graduates conducted by the institution
  D. External assessment of program quality (Professional accreditation, external review. etc.)
  E. Internal institutional assessments

Use of grant funding is not a formal selection criteria, though indication of how the proposed expenditure will enhance the program is expected, along with assurance of responsible fiscal management in a proposed budget for the award. The proposal format is limited to five pages of prose description and fifteen pages of supplementary documentation and appendices; an abstract and formatted title page are required.

For the section on program evaluation, proposal writers are asked to respond to each of he given categories but not to limit their response to those. This invitation explicitly extends flexibility to individual programs to make their best case using evidence of positive program impact on students.

Miami Selection Process

To determine its submissions for the 1985-86 state competition, Miami University established a process with considerable public visibility. A subgroup of the Council of Academic Deans wrote a description of Program Excellence and its purpose, the selection criteria, the Ohio Board of Regents review process and the University's internal selection process. This description was provided to all faculty. A timetable beginning in August and culminating in October was announced, with submissions going to OBOR by November 15 Attachment A .

The Miami proceeding for 1985-86 benefited from review of steps taken to identify submissions in 1983-84. Two years earlier, two screening committees had advised the Provost before he made selections -- one consisting solely of faculty appointees and another of academic deans. Proposals had been submitted directly to these advisory committees and the Provost. This system was adjusted to provide a single advisory committee consisting of three academic deans and three faculty members, the latter representing the three programs receiving Program Excellence awards in 1983-84. The advisory committee was chaired by a dean; it was asked to recommend six submissions in unranked order to the Provost and to provide a set of two or three unranked alternates. Proposals came to the advisory committee after divisional screening; in keeping with the University's total limit and as a further screening device, no academic division could submit more than six proposals for University consideration. Following a report by the advisory committee, the Provost made final decisions about submissions to state competition.

The advisory committee studied OBOR guidelines with considerable care before reading the eleven submissions it received from academic divisions. The committee agreed to proceed on the basis of information submitted in proposals per se rather than to solicit added information in a supplementary stage of the process before coming to judgment. The committee also felt that plans for expenditure of funds should follow logically from the claimed strengths of the undergraduate degree program, and should result in direct enrichment of that program. This view was strongly advocated by faculty members of the committee who had participated in site visits two years earlier.

In addition to its advisory role to the Provost, the committee took as an obligation and announced its intention to provide feedback and constructive criticism to any programs submitting proposals upon their request, whether or not their proposals were advanced. Evaluations of strengths and weaknesses identified in committee discussion were recorded, and shared in individual conferences between the committee chair and a representative of each proposal or the proposal writer. All programs active in the competition were encouraged to plan toward additional submissions or resubmissions in the future -- to Program Excellence. to other state award initiatives, or to external funding agencies.

The advisory committee recognized from the start that the presentation of information on student performance under the category of program evaluation would be central to the ultimate success of proposals. Notice was taken that programs are invited to expand upon the categories of information suggested by state guidelines in this area. Therefore, in its recommendations to the Provost the committee sought to make an overall judment about the strength of a particular program's evidence and did not expect uniformity in types of evidence. The key for discriminating among proposals was the relative persuasiveness of the overall individual case. In a letter of record to the Provost the committee went further:

> The Presence or absence of any singular type of qualitative or quantitative data will not unduly lend advantage to or detract from a proposal; the persuasiveness of the set of evaluative measures for a program should be the focus for judgment.

The approach taken by the Provost's advisory committee allowed for potential viability of claims based on a range of assessment initiatives in diverse programs. Proposals recommended to the Provost as six preferred submissions and two alternates included programs in Arts and Science, Fine Arts, Applied Science and Interdisciplinary Studies. Expenditure plans among finalists reflected commitments across several liberal arts fields as well as in professional programs.

Even so, greater diversity may have been desirable in submissions and perhaps in ultimate selections. In an effort to encourage this, after forwarding its recommendations the advisory committee suggested that for the next round two years hence, the time for preparation and review of proposals should be lengthened, the advisory committee established earlier and perhaps given a more leading role in encouraging proposals. The committee also sugested that assessment criteria might be elaborated more fully as a way of giving early guidance to programs.

If these steps are taken interest in assessment initiatives should expand. A workshop planned for Arts and Science department chairs in fall 1986 will seek to stimulate preparation for a third review cycle in 1987-88. In a related effort, the Provost and Council of Deans have recently opened a broad discussion of assessment approaches and their potential utility.

Assessment Mechanisms

In reviewing Program Excellence submissions sent from Miami to state competition this year, the Office of the Provost identified nine types of assessment information found in successful proposals (Attachment B). Prominent among these are traditional indicators of student achievement: post-baccalaureate awards, employment, standardized test scores, advanced degrees. Satisfaction surveys are also included; more than half of the recommended and alternate programs forwarded by the advisory committee had conducted a recent alumni survey of some kind to sample the satisfaction of graduates with their undergraduate major and their college experience.

Proposals also tended to emphasize traditional indicators of program quality, as asked for in OBOR criteria for program and resource characteristics. These included the quality of faculty as indicated by degrees, research, publication, creative achievement and service, the quality and distinctiveness of curriculum and curriculum planning, improvements in instruction design, formal program awards, results of internal and external program reviews and accreditation visits.

The Value-Added Dimension

In reporting on the first round of competition, the Board of Regents emphasized as a unique indicator of quality the concept of "value added," or "the enhancement of learner potential over and above standard expectations attributed to quality in a program." OBOR cited examples such as provision for field experiences, advanced theses and undergraduate research, special attention to undergraduate teaching and role modeling by faculty, and interdisciplinary integration of subject matter.

At Miami in 1985-86, several of the examples cited by OBOR were present in various proposals. Competitive proposals often reflected a program concept centered on teaching, advising and research with undergraduates. Evidence of high retention rates in demanding curricula was regarded as important. Information was presented showing that faculty commit personal time to individual students, direct special projects, freely give assistance to undergraduates, involve them in their own research, aid them in meeting larger career goals and in competing for post-baccalaureate educational opportunities. Examples of co-curricular activity related to academic programs in the residential university setting were not uncommon. Flexible curriculum plans encouraging opportunity to study abroad, internships and fieldwork were visible.

Effects of Local Assessment Initiatives

In evolving campus discussion during the three years that Ohio has extended opportunities for voluntary participation in Program Excellence competition, subtle shifts may be occurring in the way program quality is regarded. Underlining these is increasing interest in the effect that instruction has on undergraduate students. This cannot be attributed solely to the two rounds of competition, yet a review of proposals submitted in each suggests that faculty are making new initiatives to illustrate student accomplishment.

Relatedly, a modest broadening of purpose has begun to emerge as programs propose expenditure of award funds. The three Miami programs receiving awards in 1983-85 each bought expensive computer or laboratory equipment dedicated to undergraduate instruction. In follow-up reports this past spring each praised Program Excellence for providing resources to support teaching that otherwise were unavailable.

In this year's competition many expenditure plans feature instructional equipment purchases, and some wish to conduct research about instructional effectiveness or use consultants to improve distinctive curricula. One proposal seeks a pilot project to formally compare changes in critical thinking among sample student groups by employing a sequence of standardized tests arranged in a research design.

Even if a movement toward active sudy of student learning cannot yet be called a trend at Miami, it's plain that Program Excellence has stimulated a considerable accumulation of data showing both the academic performance of current students and various experiences of alumni. In a few instances, programs have begun or expanded data banks that allow study of information about alumni groups.

Today data collection is probably the most pervasive image associated with two rounds of Program Excellence competition. Since Program Excellence arrived almost simultaneously with a wave of information processing equipment installations at Miami, the new availability of equipment and a fresh opportunity for its application may be mutually reinforcing. Yet clear displays of data, many available for the first time, may also allow faculty and administrators to see their programs in new, altered or broadened perspective.

To the degree that perspective yields a relationship between faculty commitments to teaching and research and student learning as partly suggesed by alumni careers and achievement, appropriate steps for advancing the quality of individual programs may become increasingly apparent. To achieve broadly visible results by such a gradual process, the nurture of relative fiscal stability and a sustained state policy commitment to supplemental funding will be advantageous.

Writing about a substantial Program Excellence award to his department two years ago, one chair recently concluded:

> This award has provided an opportunity to expand our instructional capabilities in a way that never would have been possible under ordinary circumstances. It has also inspired a pride in our department and programs for our faculty and students which may be as valuable a stimulus to excellence development as the hardware itself.

A state policy that yields pride along with hardware through voluntary competition stimulating faculty initiative is likely over time to develop sustained commitments to assessing student learning. Miami University's early response with Ohio Program Excellence shows favorable signs of just this spirit.

PRINCIPAL REFERENCES

"Assessment Mechanisms -- 1985-86 Program Excellence Proposals," compiled by Roxanne Brown, Assistant to the Provost, Miami University, May, 1986, 2 pp.

"The Ohio Board of Regents Program Excellence Request for Proposals," August, 1985, 11 pp.

"Program Excellence/Miami University Institutional Guidelines." August 22. 1985, 2 pp.

Program Excellence Proposals, 1985-86, Miami University

Report on Program Exellence Process and Preliminary Results, Ohio Board of Regents, October, 1984, 29 pp.

ATTACHMENT A
Program Excellence
Miami University Institutional Guidelines

Introduction

The Ohio Board of Regents has announced the continuance of Program Excellence which is intended to recognize and reward high quality undergraduate academic programs at state-assisted colleges and universities by awarding non-renewable enrichment grants on a competitive basis.

Eligibility

The Program Excellence Advisory Committee working in conjunction with the OBOR's administration, developed the following eligibility requirements:

Programs are defined as existing academic degree programs.
Program awards are limited to undergraduate two and four year degree programs.
Programs which received awards in 1984 are not eligible for this competition.
Grants will be made on a one-time basis for purposes of enrichment, not to replace existing resources.

Proposal Contents

The proposal text has a limit of five pages, with a maximum of fiteen pages of supplementary documentation and appendices. There must be a program description including articulation of purpose, objectives, curriculum adequacy and appropriateness to purpose. A description of resource characteristics must include: student input characteristics, quality of faculty, and quality of instructional resources and supporting services. The proposal must describe internal and external measures by which the program has been evaluated, including but not limited to the following: performance of graduates on standard tests in major fields where such tests exist, follow-up evaluations of graduates conducted by the institution, external assessments of program quality, assessment of program graduates based on external measures, and internal institutional assessments. There must be a description of the amount of funds requested and how they will be used to enhance the program. The amount of the awards will range from $50,000 to $200,000. The complete guidelines from the OBOR are available in the offices of all academic deans, executive directors, and department chairs.

Internal Selection Process

Miami University may submit eight proposals with six from the Oxford Campus and one from each regional campus (no substitutions allowed). Each Dean or Executive Director of an academic division submitting proposals should establish a review procedure and must approve those proposals submitted to the Provost's Program Excellence Advisory

Committee. The maximum number of proposals that may be submitted by an academic division to the Advisory Committee is six. The Provost's Program Excellence Advisory Committee will consist of three members of the Council of Academic Deans, appointed by the President, and one faculty representative from each of the three units that won an award in 1984. The faculty representatives were named last spring.

OBOR Review Process

Proposals will be reviewed in two phases by external reviewers. The second phase will incorporate site visits. There will be one competition in this biennium with awards being made over two years.

Proposal Deadlines

August 15--RFP's and Guideline available from Deans and Executive Directors.

October 1--Deadline for proposals to Dean or Executive Director.

October 29--Deadline for proposals to Provost-fifteen (15) copies of each proposal must be submitted.

November 15--Deadlines for proposals to OBOR.

ATTACHMENT B

ASSESSMENT MECHANISMS

-- Success in competing for fellowships, scholarships, and other scholastic awards

-- Post-graduation employment statistics and types of employment

-- Quality of the universities that accept Miami graduates for advanced degree programs

-- Performance on standardized tests

-- Satisfaction surveys

-- Grade point averages

-- Number of companies coming to Miami to recruit graduates

-- Degree to which undergraduates participate in and present papers at national meetings

-- Number of graduates who go on to earn a Ph.D.

Curtis W. Ellison is Dean and Professor of the School of Interdisciplinary Studies (Western College Program), Miami University, Oxford, Ohio 45056

THE ASSESSMENT OF ACADEMIC QUALITY: A REVIEW OF EXISTING TECHNIQUES WITH EMPHASIS ON THEIR VALIDITY AND FEASIBILITY IN PUBLIC HIGHER EDUCATION

Willard F. Enteman and Pamela Irving Jackson

Assessment of the quality of academic programs and personnel is a complex process fraught with problems of reliability and validity. Valid assessment is more often based on direct observation than on analysis of symptoms. But the time and expense associated with observational processes lead legislators and higher education offices to avoid them in favor of superficial symptomatic evaluations of quality. While the latter set of techniques provide immediate, often quantitative information relatively inexpensively, they may not mirror the quality of the academic programs they are intended to assess.

This paper reviews techniques for the evaluation of the quality of academic programs and institutions. The advantages and disadvantages of each technique are considered; validity issues are addressed; and the feasibility of the technique in the circumstances of public higher education is discussed. Both direct and symptomatic evaluation techniques are considered.

I. Direct Evaluation

In general, direct evaluation of the quality of academic programs is to be preferred over symptomatic evaluation. However, in most fields investigators frequently settle for symptomatic analyses because direct ones are either impossible or expensive. This is unfortunate because there is a tendency to treat only the symptom (when symptomatic information is all that is available).

A. Evaluation by the Institution's Academic Officers and Faculty

Academic Officers and faculty at a college or university collectively have the largest body of direct information about the quality of the academic programs. They also know about the compromises in quality which have to be made in the face of scarce resources. Examples abound of such compromises.

There are a number of advantages to using the direct testimony of academic officers as a major source of information about the academic quality of the College (Bisesi, 1981). Information is readily available, it is inexpensive to collect, and it is as close to the territory as experts can get. One of the primary responsibilites of academic officers is quality control. They must insure it through the development and maintenance of strong academic offerings and through the selection and appointment of persons who are competent, who have records of achievement or who demonstrate the potential to achieve, and then through the design and administration of a system for providing rewards and means of advancement.

The testimony of academic officers and faculty members must be treated carefully and sometimes with a degree of confidentiality. Reviews, as indicated earlier, are ongoing and are aimed at improving quality and correcting weaknesses. Little is gained by exposing weaknesses in public where misinterpretation could result. Such misunderstanding could cause irreparable harm to an institution and individuals and could counter the very purpose of the quality review.

B. Evaluation by External Agents

As a result of the reliability issue, many colleges and universities have turned to external agents to do the evaluating. Since evaluation is to be done of areas which require specialized professional and scholarly expertise, no person has sufficient expertise to perform across-the-board evaluations.

An advantage of evaluation by external agents is increased reliability. In addition external agents will bring information to the institution from the "outside," and sometimes that information is either not known or not considered as seriously as it ought to be. Using external evaluators also causes a diminution of the effects of "politics" on the evaluation process. As a consequence, more professional and objective judgments may be possible.

A significant disadvantage of using external agents will be increased expenses. In addition, the external evaluators can have only partial knowledge of the institution because obviously something less than their full time attention can be given to the task. Thus, while this knowledge can be direct, there is a sense in which it can be treated as if it were in fact symptomatic only.

1. Accreditation

a. Professional Accreditation

Professional accreditation is program oriented. Some institutions have adopted a policy of insisting that wherever professional accreditation is relevant, the institution should attain it.

An advantage to professional accreditation is that it is focused. In addition, it brings experts to the campus, and it uses criterion-based standards which must be maintained. It can also support administrators in enforcing standards for faculty, student admissions, facilities, etc. Finally, professional accreditation can be a significant force in causing the allocation of appropriate resources to the program in question.

A disadvantage of using professional accreditation as a means of quality assessment is that the accreditation standards are established independently of the needs of the rest of the campus. Professional accreditation criteria are basically established by professionals in the discipline. They are responsive to the concerns of that interest group.

The professional accreditation process itself can be expensive to prepare for and to carry out. The results of the process can also be extremely expensive as the institution seeks to meet the real or imagined demands of the professional accreditation association. Consquently, professional accreditation can cause serious dislocations within a college or university's budget by placing heavy demands on resource allocations.

Finally, most of the arts and sciences programs at the undergraduate level do not have professional accreditation associations. Thus, excessive attention to the demands of professional accreditation associations can cause significant deterioration of resources for liberal education and serious conflict with the academic requirements of liberal education.

b. Regional Accreditation

Regional accreditation is the only comprehensive institutional accreditation review. Since the review is comprehensive, it allows for some sensitivity to competing demands within the institution. It has the advantage of being criterion-based, and it allows considerable latitude in institutional self-definition. It is widely accepted and virtually a necessary condition for any respectable college or university. Finally, the regional accreditation process carries within it some significant concern for general education at the baccalaureate level.

There are some disadvantages to relying on regional accreditation. The reviews may be held as infrequently as once every decade. The external team visit follows an extensive internal Self-Study, but the visit is only two and one-half days long, not much time in which to gain a solid understanding of a college or university. In addition, the process is expensive for the College to prepare for and for the Association to administer (though the team's time and efforts are donated).

2. Visiting Committees

Visiting Committees are a common source of evaluation beyond internal administrators and accreditation processes. Typically, they are oriented toward special programs, but they can be broader based (such as a school within the College or University) or even institution-wide.

They give a careful review of the program, unit or institution from an external perspective and attention can be focused clearly on specific issues and concerns. Team members may be selected because of their expertise and their interest in the college or university. They can be allowed to have a visit as long as necessary to gather the information needed to make a substantial judgment. The committee members stand outside the political process of the campus and the State and, thus, can be trusted to render more objective analyses.

If visiting committee reviews are conducted properly, they are expensive to administer. In addition, visiting teams often advocate significant changes which call for new and improved resources for the area under review.

a. *Ad Hoc* Visiting Committees

Ad hoc committees [illegible] be created to review and analyze special situations. Once such a committee has discharged its duties, it can be dismissed. An advantage to the use of *ad hoc* committees is that the focus of the committee is on the concerns of the college or university at the time it is formed. Membership can be selected with special attention to the expertise needed.

Disadvantages sometimes cited are that departments or other relevant units insist on playing such a large role in determining the membership of such committees that the credibility of the committee is brought into question. *Ad hoc* committees do not have a continuing relationship with the institution and, thus, their focus may be too limited. Given the lack of a continuing relationship, each committee must be brought up to date on the nature of the institution, its hopes and prospects, and its likely future.

b. Permanent Visiting Committees

Some of the most prestigious universities in the country have constructed a process of separate permanent visiting committees for each major program. Once initial membership for the committee is appointed, the committee itself decides upon new members and upon the terms of all members. This process takes the *ad hoc* process one step further, and gives the committee considerable independence and greater objectivity in carying out its duties.

The relative independence of the permanent visiting committee is a great strength. The committee can focus on the problems at hand without being caught up in the politics of the campus or external constituencies. Given the permanence of such committees, they begin to identify strongly not only with the programs which they are brought to review on a continuing basis, but also with the college or university which provides a home for the programs.

It is true that creating and maintaining permanent visiting committees for each major program is an expensive process because of the staffing and operating expenses necessary to engage in a significant number of reviews each year, to prepare for reviews to take place in the following year, and to respond to reviews of the prior year.

II. Symptomatic Evaluation

No one who reads current material about higher education can doubt that concerns with evaluation, assessment, and quality are dominant among professional observers of education. The impact of these concerns on the practice of teaching and learning will depend upon the extent to which their proponents ask for or supply sufficient resources to permit assessment to have an impact upon teaching and learning. Unfortunately, there is evidence that some state-wide assessment efforts are being developed on the basis of superficial and unreliable symptomatic data.

With the exception of value-added evidence, most symptomatic evidence is *norm-based*, not *criterion based*. Students and colleges are ranked in comparison to each other. Speaking logically all students or colleges cannot be in the top 50% on a norm-based comparative scale; half must be below 50%. That may appeal to the American instinct for competition and to those who want to rank colleges and universities as if they were car rental companies, but it is potentially destructive of much of the good learning, growing, and developing which goes on. Such criteria provides for only a superficial analyses of quality.

Within each category below, we point out that one can do either a superficial analysis or one of defensible quality. Needless to say the latter always takes more time and resources. That should not be surprising. Doing it right in education is expensive just as it is in medicine.

A. Reputational Rankings as Indicators

One of the common ways of evaluating the academic quality of an institution is to ask about its reputation (Braxton et. al., 1985; Struckman-Johnson, et. al., 1985; Thompson and Zumeta, 1985; Astin, 1981; Lewis and Astin, 1982).

In the words of President Bok (1985: 15) of Harvard University:

> The process of academic competition is a curious one. Athletics aside, the rankings of universities, and even professional schools, change surprisingly little over time. Virtually all of the "top twenty" institutions in 1980 were also in the top twenty in 1970, 1960, and 1950. Apparently, success attracts the abler students, better faculty, and greater funding that lead to continued success.

While there are frequent attempts to destabilize the reputational academic rankings, there is little evidence of such change. Perhaps the most popular mechanism for destabilization is the launching of powerful intercollegiate athletic programs in the so-called major male sports. While there may be some evidence that such strategies can get the public's attention, there is little evidence that the attention can be translated into respect for academic programs.

One advantage of relying on reputational indicators is that there is a presumptive correlation between reputation and quality. If a reputation is secure and stable, there may be some good reason to believe it is a valid indicator of academic quality. Even though wealthy people do spend more on everything than their poorer counterparts, they also look for value, and there is the capitalistic faith that "you get what you pay for."

A disadvantage of using reputational indicators is that reputations tend to lag behind reality and they are stable over time because of historical circumstances which may not be projectible into the future. Stanford University may be the significant exception which proves the rule of the stability of reputation. Keller (1983: 139) refers to its rise from 1955 to 1975 as one of the most dramatic success stories in higher education. But note that he says it took several big steps: "...recruit ing over 150 of America's finest minds to the faculty ...sequenced emphasis on academic areas...massive fund raising...." It can be done, but the cost is enormous. Reputations are a function largely of resources and of the wealth and social standing of students and graduates. There is no evidence that reputations are correlated with student learning.

B. Resources as Indicators of Quality

Analysis of the use and distribution of resources is frequently seen as the best symptomatic indicator of the quality of an academic institution and its programs (cf. Weathersby, 1971, 1982; Bowen, 1980, 1981; Astin, 1985). There are two forms of analysis which are done: (1) comparative and norm-based; (2) criterion and standard based.

The comparative and norm-based approach has considerable potential. However, it is very difficult to construct, and it is very difficult to gather truly comparative information. If it is done in a superficial way, it can prove to be more dangerous and destructive than other systems. In order to know that all the fiscal accounts of the institutions compared are kept in the same manner, it is necessary to have a knowledgeable authority examine the separate accounting systems from the chart of accounts up to the aggregated data.

The criterion and standard-based approach can yield interesting information, and it is arguably the best method for assessing academic quality. By way of example, consider the library. There are standards for library collections (Bonn, 1974) and standards for library budgets. Those can be examined to determine if: (1) the library collection is: (a) poor, (b) modest, (c) adequate, (d) good or (e) excellent; (2) the collection budget is keeping up in the face of the extraordinary inflation of materials, (3) the Library is able to keep up with modern technological changes, (4) the building is appropriate and prepared for future (inevitable) expansion, (5) the staff is adequately sized and compensated. There is a great deal of library information available and much of it can be used to help develop relevant criteria and standards.

Using resource analysis and criterion-based approaches, one can list accepted indicators of academic quality. In addition to the example of the library above, others would include:

a. Size of Financial-Aid Gapping

"Financial-aid gapping" indicates the amount of demonstrable student financial aid need which cannot be met by the financial aid resources. The failure to be able to meet students' needs means students who are desirable on admissions criteria may not attend the college, thus depressing the quality of the student body.

b. Faculty Salaries

The better faculty seek the higher salaries just as other people do. Better institutions generally pay higher salaries (Blau, 1973).

c. Extent of Part-Time Faculty Teaching
Most often, part-time faculty teaching is of lower quality than full-time. There are important programmatic exceptions to this rule, where, for example, particular expertise can be provided by bringing a specialist to campus on a part-time basis. And at times, enrollment projections might advise against the use of full time tenure track faculty appointments. But generally when institutions employ part-time faculty for resource rather then programmatic reasons quality falls thereby.

d. Support for Faculty Professional Development
Support for sabbaticals, research time, attendance at professional meetings, opportunities to learn about new technologies and teaching techniques is an indicator of an institution's concern for quality; lack of such support could be an indicator of lack of concern for quality or of a willingness to comprise quality.

e. Proportion of Students in Residence
The research shows a strong correlation between academic quality and the residential status of students (Alfert, 1966; Astin, 1973, 1975, 1977 Astin and Panos, 1969; Chickering, 1974; DiCesare, Sedlock and Brooks, 1972; Kramer and Kramer,1968). Support for attractive residential space increases academic quality; lack of such support reduces it. As President Bok (1982: 97) of Harvard notes, "studies have revealed that students who have the opportunity to interact closely by living together in campus dormitories show greater progress toward important educational goals than students who commute to nonresidential colleges."

f. Support for Student On-Campus Work
The research shows a positive correlation between academic quality and student on-campus work. There is a strong negative correlation between academic quality and off-campus work (Astin, 1975, 1977; Cohen, Brawer, and Conner, 1969; Kocher and Bellamy, 1969).

g. Support for Academic Supplies
Support for academic supplies, equipment and modern technology especially in the scientific and artistic areas. Faculty cannot support students' development without the necessary and proper supplies and equipment.

There are other resource indicators which have been suggested in the literature and which might be used. The advantage in using resource indicators is that they are readily available, and they give some reasonable confidence in level of academic quality. They are superior to reputational indicators though, in general, there is a relationship between the two.

The disadvantage in using resource indicators is that they do not measure how much (or little) students have learned. They are "input" indicators. It is possible to imagine a high level of resources and little impact on academic quality. Thus, while resource indicators are reasonably good surrogate indicators, one would want some assessment of student learning for a more complete picture.

C. Outcomes

The interest in "outcomes" information as a measure of academic quality arose in large measure in response to the dissatisfaction with reputational hierarchies, especially those built primarily on resource bases.

As pointed out earlier, attention to resource issues is focused on "input," so it is not surprising that the reaction of higher education scholars would be to begin to argue for concentration on "outcomes." Weathersby (1971) and others began to question seriously the validity of input measures, and argued, rather, that evaluations should be established on the basis of student results.

The outcome measures suggested include indicators such as student performance on Graduate Record Examinations, on Law School Aptitude Tests, on Medical College Aptitude Tests, on Graduate Management Achievement Tests, on student placement in "prestigious" graduate schools, etc. There is a high correlation between scores on exit multiple choice examinations such as the GRE and scores on entrance exams such as the SAT. The highest nonacademic correlates of the SAT are family income, background and number of siblings. Students who come from poorer families and students who have more brothers and sisters tend to have lower entrance examination scores. Thus, the prestigious institutions found that outcome measures rank in their favor (cf. Astin, 1972; Karabel, Jerome, and Astin, 1975).

Outcome measures return one's attention to educational achievement. However, they may not measure educational growth or achievement in general, and they certainly do not measure educational growth and achievement as a result of the student's college education. All that outcome measures can point to is the level of performance of some students at a certain age. The unexamined question is whether the level of performance is an "outcome" of the college education or whether it is incidental to the college education.

Given the fact that high SAT scores predict high GRE scores, there is some reason to question how much gain is a result of the college or university experience. The fact that all of these scores are norm-based--the GRE scores providing comparisons among college graduates and the SAT scores providing comparisons among high school seniors--makes it difficult to assess student change scores in light of this question. Criteria-based outcome measures would provide a clearer picture of college impact on students' growth.

The disadvantage of outcome measures is that they can be used and abused on purely symptomatic grounds. It is not hard to discover how to move these numbers around by internal institutional activity.

D. Value Added

Almost surely, Astin is the intellectual father of this approach, though references to its appropriate analogy show up in advanced management literature (cf. Drucker, 1939, 1954). It is instructive to note that the American Assembly of Collegiate Schools of Business was perhaps the earliest accreditation organization to contemplate using value added concepts in the standards.

The psychologist, Dr. Kenneth Clark, known for his expert testimony in the Brown vs. the Board decision, and soon to be one of the ten medal recipients at the July 4 Statue of Liberty ceremonies, has recently attacked punitive and rejective uses of Student Outcome Assessments. He has argued that by their nature such uses reinforce current patterns of stratifaction in education by physically or psychologically destroying some students at a vulnerable stage of intellectual development. Instead, he stresses, that the purpose of testing should be diagnostic. Evaluations should help identify a student's growth and the system that maximizes student development. Failing to do that means one has made a decision on the basis of questionable evidence that some people are simply uneducable.

The value added approach seeks to identify where students are when they enter the college, where they are when they graduate and, if possible, where they are later. Its ideal is to also control for growth attributable to causes other than College, thus identifying the growth which is attributable to the College experience. The task is a difficult one and its completion will require longitudinal data and creative social scientists. Successful development of student assessment data and use of the data in curricular revisions is expensive. It requires the involvement and acceptance of the academic officers, faculty, and students of an institution and of a multitude of data sources. Finally, the process requires patience. Northeast Missouri State, one of the two colleges to have successfully implemented a value added program, phased it in over a decade (cf. Astin, 1985, 1982; Whitla, 1981, 1977; Northeast Missouri State University, 1984).

The value added approach has a significant advantage over other indirect and even over the direct approaches to assessing academic quality in that it is the only one which attempts to measure academic quality by assessing the educational effectiveness of the College on students.

With some development, it can also raise other questions about resources. For example: do incremental resources produce improved educational effectiveness? Is there a diminishing returns curve with regard to resource commitment? Is one system of teaching more effective for students than another? Do smaller classes yield greater educational growth? Are some curricular paths demonstrably better than others in promoting the growth of some students? And so on (cf. Weathersby, 1977; Romney, 1978).

There are some disadvantages to the value added approach. It is expensive to start and maintain. Once started, a great deal is lost if it is stopped. There is a substantial lag time between the initiation of the process and the arrival of usable data. Even for students who complete their degree in four years, it will be four years before any post-test data can be discovered. In addition, pre-test/post-test collegiate data does not give an assessment of later accomplishments. The graduate may not bloom until some years after graduation, but the conditions for growth may have been well established in the person's college career. Once this extensive date is collected during the students' college years, the institution should attempt to keep it alive.

REFERENCES

Alfert E., 1966, "Housing Selection, Need Satisfaction, and Dropout from College." Psychological Reports, 19: 183-186.

Astin, A.W., 1985, Achieving Educational Excellence. San Francisco: Jossey-Bass.

1985, "Involvement: The Cornerstone of Excellence" Change, v17 July-August 1985

1982, "Why Not Try Some New Ways Of Measuring Quality?' Educational Record v.63 Spring, 1982

1977, Four Critical Years. San Francisco: Jossey-Bass.

1975, Preventing Students From Dropping Out. San Francisco: Jossey-Bass.

1973, "Impact of Dormitory Living on Students." Educational Record, 54: 204-210.

Astin, A.W. and R.J. Panos, 1969, The Educational and Vocational Development of College Students. Washington. D.C.: American Council on Education.

Bisesi, Michael, 1984, "Program Evaluation, A Qualitative Planning Tool. Planning and Changing: volume 14, number 3: 144-151.

Blau, Peter, 1973, The Organization of Academic Work. New York: John Wiley and Sons.

Bowen, H.R., 1981, "Cost Differences: The Amazing Disparity Among Institutions of Higher Education in Educational Costs Per Students." Change, volume 13, number 7: 14-19.

1980, The Costs of Higher Education: How Much Do Colleges and Universities Spend Per Student and How Much Should They Spend? San Francisco: Jossey-Bass.

Bok, Derek, 1982. Beyond the Ivory Tower. Cambridge: Harvard University Press.

1985, "The Quality of Education in American Universities" The Centennial Celebration of the New England Association of Schools and Colleges

Bonn, George S., 1974, "Evaluation of the Collection," Library Trends January

Braxton, John M. and R.C. Nordall, 1985, "Selective Liberal Arts Colleges: Higher Quality as Well as Higher Prestige?" Journal of Higher Education, Volume 56, number 5: 538-54.

Business Week, "Every Campus for Itself" March 24, 1986 Number 2938

Chickering, A.W., 1974, Commuters Versus Residents. San Francisco: Jossey-Bass.

Cohen, A.M., F.B. Brawer, and A. Connor, 1969, Selected Personality Correlates of Junior College Dropouts and Persisters. Paper read at California Educational Research Association. Los Angeles."

DiCesare, A.C., W.E. Sedlacek, and G.C. Brooks, Jr., 1972, "Nonintellectual Correlates of Black Student Attrition." The Journal of College Student Personnel, 13: 319-324.

Drucker, Peter, 1939, The End of Economic Man John Day Co.
1954, The Practice of Management, Harper Row

Keller, George, Academic Strategy, 1983, The Johns Hopkins University Press

Kosber, E. W. and R.O. Bellamy, 1969, "Some Background Differences Between College, Graduates and Dropouts." Psychology, 6: 1-6.

Northeast Missouri State University, 1984, In Pursuit of Degrees with Integrity. Washington, D.C.: American Association of State Colleges and Universities.

Potts, David, 1971, "American Colleges in the Nineteenth Century: From Localism to Denominationalism" History of Education Quarterly v. 11

Romney, Leonard, 1978, "Measures of Institutional Goal Achievement" WICHE Contract

Rudolph, Fredrick, 1965, The American College and University: A History, Random House

Struckman-Johnson, Cindy and Kinsley, Steven, 1985, "Assessment and Comparison of College Image among High School Seniors, College Students, and Alumni" College and University v. 60 Summer

Thompson, Fred and Zumeta, William, 1985, "Hiring Decisions in Organized Anarchies: More Evidence on Entrance into the Academic Career" Review of Higher Education v. 8 Winter

Weathersby, George B., 1971, "Input, Output, and Institutional Objectives" Compact v.5 April

Whitla, Dean K., 1977, "Value Added: Measuring the Impact of Undergraduate Education" FIPSE

1981, "Value Added and Other Related Matters" Department of Education

Willard F. Enteman is Provost and Vice President for Academic Affairs at Rhode Island College, Providence, R.I. 02908

Pamela Irving Jackson is Chairperson and Professor of Sociology at Rhode Island College, Providence, R.I. 02908

WHERE IS THE VALUE IN 'VALUE-ADDED' TESTING?

Howard T. Everson

At the insistence of legislatures and state boards of higher education, an ever increasing number of public colleges and universities are developing and using "value-added" assessments in an effort to measure learning in college. In general, value-added assessments compare students' achievement in academic subjects and in such "life skills" as analyzing a newspaper column, a mathematical table, or T.V. ad using a pre-post design, i.e., before and after a prescribed period of college study (Jaschik, 1985). In value-added terms, an "excellent" institution fosters significant improvement in the cognitive and affective functioning of its students.

Supporters of the value-added approach claim that the information obtained from the tests permit more accurate curricular refinement and enhanced remediation. Moreover, they point out that students also receive concrete measures of their educational attainments. In contrast, critics have voiced the concern that value-added testing may foster uniformity of curricula. And further, the testing program may serve to force the faculty to teach only those subjects covered on the standardized, summative exams.

This paper outlines the basics of value-added testing programs and discusses the psychometric design issues they raise. The essay concludes by presenting an alternative design for assessing more plausible models of educational achievement that could be used by many colleges and universities without incurring the enormous start-up costs associated with mass testing programs.

The causal model outlined in this paper holds the promise of providing both administrators and researchers with a robust method for examining the effects of higher education, in terms of learning and achievement, without incurring the enormous start-up costs of a comprehensive testing program. Models such as the one proposed here, as well as others (see Wahlberg & Shanahan, 1983), have been used with success to examine the effects of secondary education on a variety of populations. Coupled with a recently developed class of measurement techniques, referred to as linear structural equations or LISREL (Joreskog & Sorbom, 1979), completely specified causal models of educational achievement can provide a powerful, cost-effective method for estimating the effects of an array of colleges, curricula, and programs on a variety of student populations. Moreover, these modelling techniques have the powerful, cost-effective method for estimating the effects of an array of colleges, curricula, and programs on a variety of student populations. Moreover, these modelling techniques have the powerful advantage of providing comparable achievement models for diverse groups of students. Keeping in mind the divergent groups of students served by large, urgan universities, that advantage is highly desirable.

National Perspective

Value-added testing differs from the competency tests currently in place at colleges and universities around the country. Unlike competency tests which focus on reading, writing, and mathematics, value-added tests attempt to determine how the college experience has raised students' achievement in academic areas and improved their ability to function in society. Although some colleges use value-added tests to advise students, they have not required students to score certain grades to advance to a higher level of education. State coordinating boards require value-added testing in Tennessee and South Dakota. Several other states, including Colorado, Maryland, New Jersey, and Virginia, are considering similar requirements. Individual institutions, public and private, ranging from large research universities to small, non-traditional colleges also have started such testing.

Tennessee's Higher Education Commission requires universities to have self-evaluation programs to qualify for "performance-based funding," a percentage increase in their budgets. The Commission does not require all institutions to use a single test; instead it requires them to report on how they are measuring students' learning. The University of Tennessee at Knoxville, for example, uses the American College Testing College Outcome Measures Program, which is designed to measure the effect of general education on increasing students' ability to analyze writing, use science and technology, and understand the workings of social instituions. Tennessee's students take the test first as freshman and again as seniors, and the change in scores is used to analyze the effectiveness of the curriculum.

In South Dakota, half of all freshman at public institutions were tested this past year as the first part of a value-added program ordered by that state's Board of Regents. Eventually the Regents plan to have all the students tested as freshmen and at the end of the sophomore year, using the A.C.T.'s test of outcomes. Seniors will also be tested, using a standardized examination in their major fields. South Dakota's program was developed after state higher education officials visitied Northeast Missouri State University, where value-added testing has been in use since 1973. At Northeast Missouri, students take the A.C.T. outcomes test of the A.C.T. college entrance exam when they enter as freshman and at the end of the sophomore year. Seniors at Northeast Missouri take a national standardized test in their major subject.

In an effort to determine the contribution colleges make to the citizens of New Jersey, that state's Chancellor for Higher Education, T. E. Hollander, has proposed the College Outcomes Evaluation Program. This program calls for a comprehensive evaluation of the outcomes of higher education. The "outcomes" are clustered into three areas: (1) student performance, as a measure of individual achievement and as an institutional outcome; (2) community-based institutional outcomes; and (3) society-wide institutional outcomes. Although the program's scope is comprehensive and will attempt to assess community and society-wide effect (e.g., employment patterns, cultural developments, and the development of new knowledge), central to New Jersey's plan is the notion that the outcomes or effects of higher education can be measured by assessing the intellectual development of individual students from a "value-added" perspective.

Psychometric Issues

In general, value added testing programs call for assessing performance at three points during students' college careers: (1) at entry - this take the form of freshman skills assessment and, as such, provides a baseline measure of the basic skills for the "value-added" analysis; (2) at midpoint - viewed as the "rising junior" level, or entry into the upper division or transfer from two- to four-year college; and (3) at graduation - this culmination point establishes the minimum standards for receipt of the B.A. degree and completes the "value-added" equation. Moreover, at each level, the plan calls for assessment of several different kinds of learning. Among them are basic skills in reading, writing, and mathematics; cognitive development or higher order thinking ad reasoning; general educational development, usually viewed as a combination of basic and reasoning skills; and domain-specific knowledge in the student's area of specialization.

Leaving aside for the moment all the obvious problems inherent in the development, standardization, and implementation of such a comprehensive testing program, a more subtle, but no less thorny, problem with this program has to do with the causal model of educational attainment implicit in the typical value-added plan. The specification of a valid causal model of educational achievement, particularly in higher education, is not a simple matter. A misspecified achievement model, for example, can invalidate conclusions of the effects of higher education on academic achievement.

To some observers of higher education, the elaborate methods typically presented to assess higher education outcomes seem far too complicated and unnecessary. Why, critics ask, could we not simply project an individual's intellectual growth and then see whether going to college stunts or accelerates the growth curve? Proponents of this approach suggest that if we simply gather pretest and posttest achievement data on students enrolled at a particular college or university-level program, the observed gains will reflect both natural intellectual growth, as well as the effects of attending college. Measuring higher education's effect, then, is simply a matter of separating out those components of intellectual growth.

The "value-added" approach, however, as it is presented in the educational literature (see Bryk, et al., 1980), was developed around the concept of an individual growth curve. For example, over the time from pretest to posttest it is assumed that a student's intellectual growth consists of two components: (1) systematic growth, which can be characterized by a growth rate and a corresponding growth curve; and

(2) individual variability, a random component which is specific to a particular student at a certain time.

In contrast to studies which attempt to describe achievement levels in various educational sectors, value-added testing programs are, in fact, explanatory studies. That is, their aim is to explain why students in some colleges achieve more than others by attempting to separate the effects of attending college from the natural intellectual growth of the individual. As an explanatory model, then, such a testing program implies a proposed causal structure that includes both well-established and hypothesized causal relationships. Value-added testing programs attempt to explain achievement as a function of basic academic ability and intellectual growth while in college. Figure 1, below, presents the essential structure of the causal model implied in the typical value-added testing program.

The "value-added" model clearly implies causal relationships between entering level basic skills, intellectual growth, and educational goal attainment. There are obvious problems with the causal relationships specified by this model, not the least of which is the absence of background measures of motivation, socio-economic status and cognitive ability. Consequently, and perhaps more seriously, the model begs the question of whether intelligence is a unitary construct that can be captured by current measures of basic skills, problem solving ability, and subject matter knowledge. These are serious theoretical issues that, if ignored, present strong threats to the validity of the conclusions reached by those doing the evaluating. Consequently, if the goal is to provide colleges and faculty with useful data for reshaping programs and curricula, the implementation of an incomplete causal model of achievement may produce a disappointingly low yield.

As noted earlier, the "value-added" model assumes that we can reasonably approximate an intellectual growth curve for each student. To plot a growth curve, however, we need to have the same units of measurement across time. For example, for height, inches or centimeters would be used. In educational research, however, the choice of the units of measurement is not so clear. For variables in which students are expected to grow (e.g., skill aquisition, critical thinking, and knowlege), different tests are needed for different ages or else serious ceiling or floor effects could result. Calibrating the units of measurement of different tests is not a simple problem. Even if exactly the same test were used as pre- and post measures, spurious "growth" would result due to the uncontrolled influence of individual differences in background, motivation, maturation and test-taking ability.

From the standpoint of learning theory, the "value-added" model as presently construed presents the researcher with some serious design flaws which threaten the validity of its conclusions. By striving to reduce a complex psycho-educational phenomenon into a simple pre-post gain score equation, the "value-added" plan presents a theoretical model of intellectual growth which is simplistic and counterintuitive. The research literature in psychology and education clearly suggests that

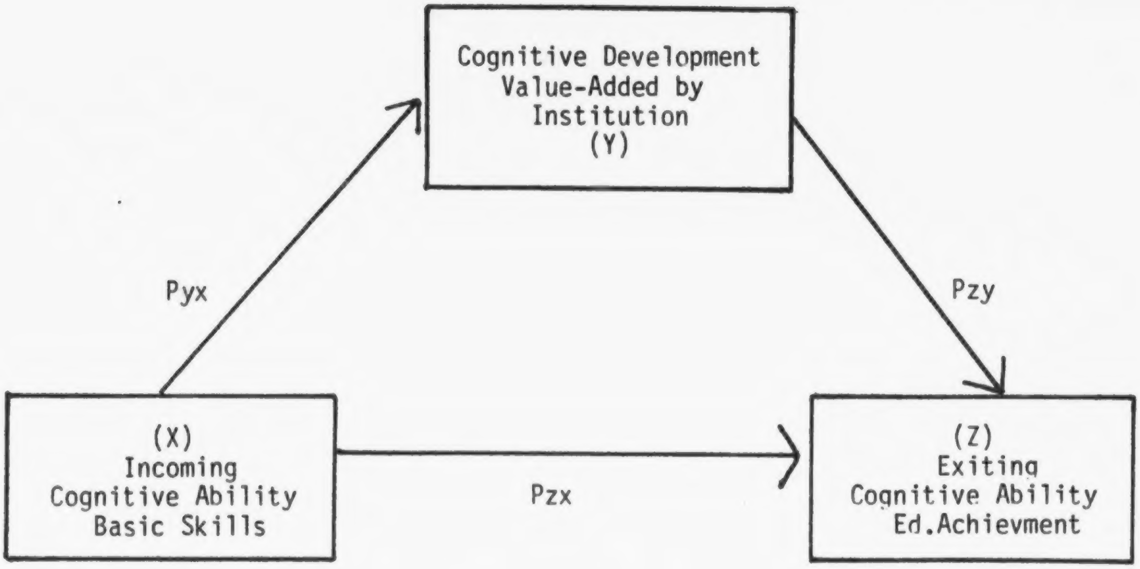

Cognitive Development
Value-Added by
Institution
(Y)
Pyx
Pzy
(X)
Incoming
Cognitive Ability
Basic Skills
Pzx
(Z)
Exiting
Cognitive Ability
Ed.Achievment

Figure 1: Simple recursive 'value-added' model of educational achievement

academic achievement, whether it is on the primary, secondary, or postsecondary level, is a complex function of economic status, motivation, ability, and educational experience, and not simply a function of whether or not one attends one college or another (Wahlberg & Shanahan 1983; Wolfle, 1985).

Second, the bulk of the "value-added" models developed in the technical literature are deterministic. That is, time brings about patterned changes that may vary across individuals but, nonetheless, are driven by time. Although such models may provide useful estimates of physical growth, where genetic and environmental components are better understood, intellectual growth does not necessarily follow a linear or deterministic path and assessment methods must recognize its multidimensional nature. Unavoidably, and to the dismay of legislators and other interested observers of education, modelling intellectual growth and achievement becomes an increasingly complex process.

From a research perspective, then, what is needed is the development of plausible causal models of educational attainment. At the very least, these models must include measures of individual differences in ability, motivation, and socio-economic status, along with measures of the type of curriculum and type of college.

An Alternative Model

The ambitious scope of the value-added plan is commendable. However, the danger inherent in such a comprehensive initiative is that a complete model of the causes accounting for educational attainment will go undeveloped. The mandate to researchers, then, is to develop and test convincing, plausible models of academic achievement that yield informative, useful data on the effects of higher education.

In general, a causal model consists of the theoretical relationships among unobservable or latent variable. Examples of reasonable, yet parsimonious, causal models of educational attainment are readily found in the research literature. For example, as an alternative to the "value-added" model, the causal model presented by Wolfle (1985) may prove useful and easily adaptable. Wolfle's model considers educational achievement to be a function of socio-economic status, ability, college curricula and college grades. This model, presented in Figure 2 below, depicts the latent variables by elipses, while the observable measures are indicated by rectangles.

Measurement of the unobservable variables in this alternative design requires collecting data on two or more indicators of each latent variable in the causal model. Multiple observable measures permit the estimation of variances and covariances among the latent variables of interest and, thereby, provide reasonable estimates of the relative strength of the causal relationships among them.

Socio-economic Status - This variable can be measured using relatively straightforward indices of family income and parents' educational status.

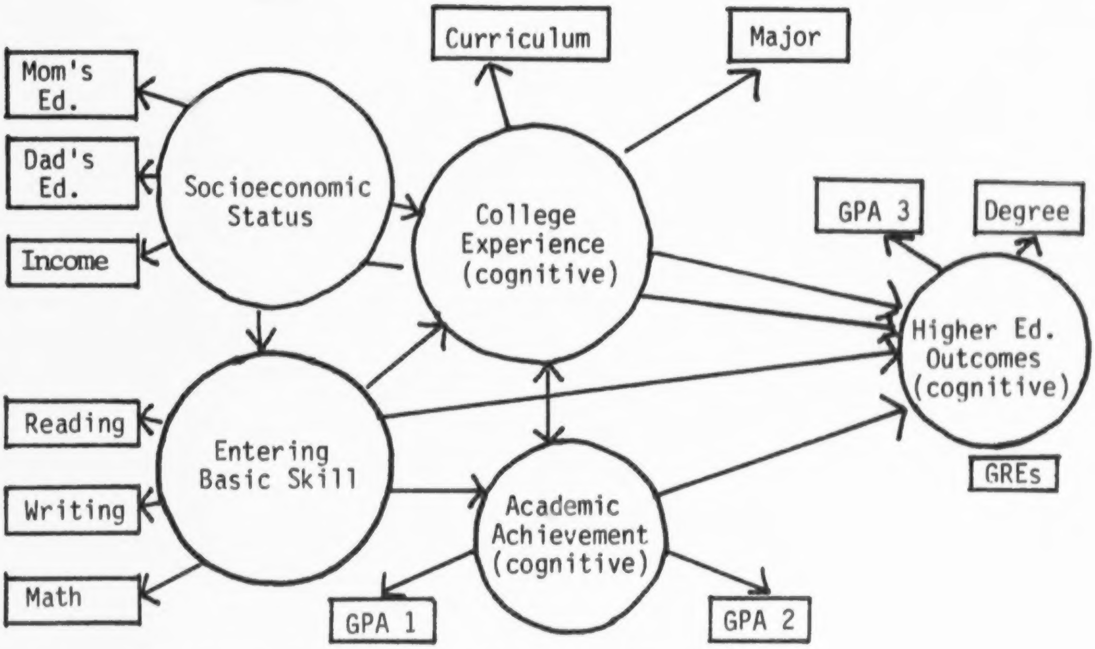

Mom's Ed.
Dad's Ed.
Income
Socioeconomic Status
Curriculum
Major
College Experience (cognitive)
GPA 3
Degree
Higher Ed. Outcomes (cognitive)
GREs
Reading
Writing
Math
Entering Basic Skill
Academic Achievement (cognitive)
GPA 1
GPA 2

Figure 2: LISREL Model of educational achievement.

Ability - This variable can be measured using the three scores (i.e., math, reading, and writing) obtained by during entrance testing or other competency testing programs.

College Curriculum - This can be measured by reports from the participants and the colleges. It would be of interest, for example, to know the comparative effect on academic achievement of attending a college with a core curriculum in place versus attending a college without such a requirement.

College Grades - Again, this variable would attempt to measure achievement at some point during the college experience, say, after completion of the first 30 credits and again upon completion of 65 credits.

Educational Attainment - This variable could be measured by gathering data on whether the participants earned a B.A. degree, number of credits earned, cumulative GPA and GRE scores.

The causal model outlines above provides researchers with a method for examining the effects of higher education without incurring the enormous start-up costs of a comprehensive testing program. Models such as the one proposed here, as well as others (see Wahlberg & Shanahan, 1983), have been used with success to examine the effects of secondary education on a variety of populations. Coupled with a recently developed class of measurement techniques, referred to as linear structural equations (LISREL), completely specified causal models of educational achievement can provide a powerful, cost-effective method for estimating the effects of an array of colleges, curricula, and programs on a variety of student populations. These modelling techniques have the powerful advantage of providing comparable achievement models for diverse groups of students. Keeping in mind the divergent groups of students served by large, urban universities, that advantage is highly desirable.

Recommendation

Clearly, we must produce and test more convincing causal models of educational attainment than the one offered as "value-added." In considering possible models, it may be useful to distinguish between exploratory and more "confirmatory" stages in the investigation of a phenomenon. Exploratory approaches are particularly useful in current studies of the effects of educational programs because of the primitive state of relevant theoretical models. At the exploratory stage, the data analysis should suggest ways in which the theoretical model might be modified, how the measures might be combined or separated, or which variables can safely be deleted from the model.

The causal model adapted from Wolfle (1985) lends itself to pilot testing. The obvious advantage of such a pilot study is that it can be implemented with only a fraction of the costs, both political and financial, of mounting a comprehensive testing program. Moreover, small scale evaluative research can provide invaluable data on how well various achievement models fit diverse student populations. Such estimates must precede the implementation of a costly, comprehensive testing program.

Assessment in higher education is nothing new. Although educational measurement is an imperfect science and the application of measurement results in a nebulous art, it is not as though the academy is starting from scratch. Good examples of instructional approaches that integrate assessment with undergraduate instruction exist. Among them are the Hutchings College at the University of Chicago, the General College of the University of Minnesota, and several dozen smaller liberal arts colleges around the country. Comprehensive testing at several levels, the use of outside examiners for departmental majors, and the administration of such standardized tests as the GRE to graduating seniors all are part of such curricula. In addition to pilot testing causal models of achievement, the programs currently in place around the country may serve as useful implementation models and as such should be thoroughly investigated for their potential use and adptability to other institutions.

References

Bryk, A., Strenio, J. & Weisberg, H. (1980). A method for estimating treatment effects when individuals are growing. Journal of Educational Statistics. Vol. 5, 1, 5-34.

Ewell, P. (1985). Assessment. What's it all about? Change. November/December, 32-36.

Jaschik, S. (1985). Public universities trying tests and surveys to measure what students learn. Chronical of Higher Education, Vol. XXXI, No. 3.

Joreskog, K.G., & Sorbom, D. (1979). Advances in factor analysis and structural equation models.. Cambridge, MA: Abt Books.

Turnbull, W. Are they learning anything in college? Change. November/December, 23-26.

Whalberg, H. (1984). Improving the productivity of America's schools. Educational Leadership. Vol. 41, 8, 19-30.

Wahlberg, H. & Shanahan, T. (1983). High school effects on individual students. Educational Researcher. Vol. 12, 7, 4-9.

Wolfle, L. (1985). Postsecondary educational attainment among blacks and whites. American Educational Research Journal. Vol. 22, 4, 501-525.

Howard T. Everson is Director of Testing and Measurement at the City University of New York, 535 E. 80th Street, NYC, NY 10021.

AN APPROACH TO "IN-HOUSE" TEST DEVELOPMENT FOR MAJOR FIELD ASSESSMENT

Anthony J. Golden

Tennessee, like many other states, requires that its colleges and universities assess their graduating seniors to insure that they have acquired the requisite knowledge in their major field. Faced with having to test numerous disciplines either without commercial tests being available or using expensive nationally normed tests, many institutions have developed their own assessment instruments. Instruments developed in this fashion often suffer from numerous shortcomings not the least of which is that their norms are internal and do not demonstrate their students' proficiency compared to students at other institutions.

When faced with meeting requirements for assessment of their graduates, heads of several Psychology departments in Tennessee met in 1983 to form a consortium for this purpose. A plan was developed which is now being used by numerous Psychology, Social Work, Philosophy and Geology departments throughout Tennessee. The tests are constructed from items randomly selected from jointly contributed question pools. Once taken, the tests are scored together and each institution's scores are determined using a composite norm. A number of features have been included in order to permit departments to use their test results not only to meet state requirements but also for internal program evaluation.

This paper describes several techniques which can be used for test construction and evaluation and the methods which are available for providing useful feedback on test performance.

An important difference exists between course oriented examinations and terminal proficiency tests. Course examinations must be designed to assess mastery of specific material covered by the objectives of the course. Proficiency tests must cover a much broader amount of material, assuming mastery and the ability to integrate information from a number of different areas. Course examinations must be reasonably reliable and content valid. Proficiency tests often must meet the additional criteria of predictive and construct validity. The items used in such an instrument must be appropriate to the discipline being tested, conform to an established definition of terminal proficiency and, ideally, be able to predict future performance in the graduate or professional setting. The test itself must conform to an overall model of the "ideal" graduating major. Whereas a student with a 36 semester hour major field concentration will take 12 final examinations consisting of 100 or more multiple choice items or their equivalent in essays and practical demonstrations, a proficiency test must cover a similar amount of material, at a level of terminal mastery, using only 200 to 300 items. Obviously, both the test and the items used to construct the test must be designed to assess the critical aspects of the field in as compact and concise a manner as possible while maintaining a high degree of overall validity.

Developing the Testing Model

The first step in the construction of an "in-house" proficiency test is to determine the areas that it will include and their relative contributions to the overall score. This paper discusses three different approaches for making decisions of this nature. The first simply is to obtain a consensus from the department constructing the test or from the group of departments which make up a testing consortium. An outline is constructed and agreed upon which lists the basic areas to be covered and the specific points to be tested in each area. Equal weighting is then given to each area with the questions within each area being divided equally among the points to be tested. Several problems exist within this approach, however. It is often difficult to obtain a working consensus of what areas are and are not important within an academic discipline. The result can all too easily be one of too many areas to test and not enough time or questions to test them adequately. In addition, most academic disciplines are not evenly divided amongst their various specialty fields, all departments do not offer all fields, and most students will not be exposed to every course offered by a department in their major field.

An alternate approach is to balance the examination using a statistical sampling technique. The department or departments agree upon up to 10 discrete specialty fields for inclusion in the test. Typically these are fields that are required or strongly recommended for completion of a major. Each field is then subdivided into specific points to be tested resulting in an outline of the examination contents. Once these aspects are agreed upon, the department or departments contribute questions to a common pool. Each question is classified according to which aspect of the outline it measures. While each specialty field in the outline is given an equal number of questions, the actual number of questions to be assigned to each point within each field is based upon the proportion of questions available within the pool. The assumption here is that more frequently tested points will result in more questions in the pool. This saves the test designers from having to make decisions about including some emphases and ignoring others. Once again, several problems arise as a result of this approach. Agreement must be reached as to which fields are to be included. Within a single department, this may be relatively easy. Within a consortium of departments, however, there may be complete agreement on some fields and disagreement on others. With a large enough consortium, this problem can be solved by having one or two optional fields available to each participating department with the remainder of the test consisting of a common core of fields.

The role delineation approach to proficiency test construction can be quite time consuming although it may well provide the most balanced instrument for educational outcome assessment. At the outset, a representative group of educators in the area of study to be assessed must construct a general statement of the optimum parameters of terminal proficiency in the academic discipline. Additional statements are then written so as to specify individual competencies in those areas considered to be critical to the overall definition of proficiency.

Having described the requisite areas of education, it is necessary precisely to specify within each area those facts and skills which are considered essential. These specifications should be written in the form of individual behavioral objectives stating operationally what the graduating student should know and be able to do. These objectives must be stated in a manner specific enough so as to permit direct assessment and to provide a clear distinction as to whether or not the student has met the objective. Each objective must deal with a single aspect or a cohesive group of aspects of the overall area and must define precisely what the examinee should know or be able to do in order to be considered knowledgeable about the material.

The following is an example of the statement of a basic competency with accompanying behavioral objectives.

Learning Disabilities - Competency I

Definition/Basic Terminology: The individual knowledgeable in the area of learning disabilities shall be able to define the term "learning disability", differentiating this disability from other learning difficulties.

The individual must be able to:

1. summarize the definition of learning disabilities as stated by the National Advisory Committee on Handicapped Children.

2. state why we generally have come to accept the term "learning disabilities" rather than such terms as "minimal brain dysfunction" or "minimal neurological impairment..."

Once they are completed, the basic competencies must be ranked by order of importance. Numerous techniques are available for ranking these statements with the most straightforward being simply to rank the competencies from 1 to n with 1 being most important and so on. The major problem with this approach is that, with large numbers of statements to rank, it can become unwieldy and confusing. A more effective and less confusing alternative is to order the competencies by pairwise comparison. The statements are examined in pairs with the more important of the two being designated for each pair. The results of these comparisons can be used to construct an accurate ranking of the overall set of competencies and may provide finer levels of discrimination than the simpler linear ranking approach. An example of the outcome of a pairwise comparison among three competencies is given below.

	More Important		
	I	II	III
I	-	7	2
II	3	-	2
III	8	8	-
Totals:	11	15	4
Proportion:	.37	.50	.13

Once this initial ordering of competencies has been accomplished, the relative contribution of the objectives within each competency must be determined. Once again, a linear ranking approach can be used but a better technique presents itself in the use of a seven point scale. The scale, anchored only at the extremes, is used to indicate the perceived importance of each behavioral objective. The table below illustrates the possible outcome of such a rating.

	Minimal				Maximal			Mean	%
	1	2	3	4	5	6	7		
I.1		3	5	2				3.9	54.2
I.2	3	2	4	1				3.3	45.8
II.1			2	5	2	1		5.2	47.7
II.2				5	3	2		5.7	52.3
III.1		1	2	4	2	1		5.0	38.5
III.2		2	4	3	1			4.3	33.0
III.3	1	1	2	3	2	1		3.7	28.5

Using the basic competencies ranked by pairwise comparison and behavioral objectives ranked using a seven point scale permits a composite model to be constructed of the "ideal" graduating senior which can be translated directly into a standardized measurement instrument. The outline below is based upon a hypothetical 100 item test using the proportions developed in the previous examples.

I. Definition/Basic Terminology
    A. NACHC definition (20 items)
    B. Exclusion of other terms (17 items)

II. Symptoms of Learning Disabilities
    A. Most frequent common symptoms (24 items)
    B. Symptoms by type of disability (26 items)

III. Basic Diagnosis and Early Remediation
    A. Problems and benefits (5 items)
    B. Mainstreaming vs. self-contained classroom (4 items)
    C. Tennessee Law 839 (4 items)

The role delineation approach reduces the problems with the consensual approaches mentioned earlier and provides a more statistically valid framework for test construction.

Developing Test Items

Once a set of behavioral objectives has been constructed, it will serve as the basis for construction of the measurement instrument. Test items intended for use on such as instrument specifically should be designed to measure the abilities stated in the objectives and must be keyed to the objectives that they represent so that composite area scores can be developed.

Bloom et al (1981) specify six taxonomic levels of increasing complexity at which learning objectives may be structured in the cognitive domain. These are knowledge (level 1), comprehension (level 2), application (level 3), analysis (level 4), synthesis (level 5), and evaluation (level 6). Each of these levels has with it a corresponding level of test item complexity. For the most part, multiple choice test items fall within the first 3 or 4 levels while essay questions frequently operate at levels 5 and 6. It is possible, however, to write multiple choice test items which operate at any one of the 6 levels. When constructing test items which are intended to assess specific objectives, it is necessary first to determine the taxonomic level of the objective. The complexity of the objective should then be used to determine the appropriate level of difficulty of the test items by which it will be assessed. In most cases, test items should be constructed at the taxonomic level of the objective as well as at lower levels. Approaching the initial estimation of test item difficulty in this manner facilitates a match between the intent of an objective and its measurement.

Test items used for major field proficiency testing must assume that the examinee has been exposed to a wide variety of material in a number of different courses. For test items which are restricted to a single objective, the body or stem of the question must carefully limit the material being dealt with while the alternatives should not provide distractors which are, or can be, correct from the perspective of a different area of study. Four or five alternative multiple ,choice items should be used and should be structured so as to avoid negative" phraseology, "all of the above" and "none of the above" type choices. The alternatives should be succinct and to the point and should not distract from or distort the actual question contained in the stem. At least two (for 4 choice) or three (for 5 choice) of the incorrect alternatives should be plausible distractors. That is to say, they should be such that the examinee who is confused about the material or who has mastered it weakly or not at all is more likely to select an incorrect alternative than to guess correctly. Commonly made mistakes or misinterpretations should be used whenever possible as plausible distractors. Single test items can also be constructed so as to assess multiple objectives by making the stem require the examinee to apply information from several areas in order to seek a correct answer. This type of question contributes significantly to increasing the content validity of the instrument without adding to its length.

Test Construction

When the actual test is constructed, sufficient questions should be available to adequately measure all aspects of each objective although only a random sample of the items will actually appear on any given form of the test. The test is constructed from the item pool in accordance with the outline which has been prepared. Whenever possible, multiple forms of the examination should be prepared. The use of multiple forms serves two distinct purposes. The first is to reduce the chances of cheating. The second is to double the number of items upon which validating statistics are collected on each administration. When the sample to be tested is too small to permit multiple forms using different questions, it is still possible to construct different versions of the same form. The easiest method for doing this is to randomly rearrange the alternatives under each question. If the forms are distributed in alternating order, this will make it all but impossible for two students sitting side by side to determine correct answers from each other although they are answering the same questions.

The length of the test must be determined in order to conform to both the amount of time which will be allowed for its completion and for the number of competencies which must be covered. A minimum of 10 items should be provided on each competency with 15 to 20 being a better number. Students using computer-assisted autotutorials at Austin Peay select questions at a rate of 1.6 per minute. This would translate into a rate of 96 questions per hour. Allowing for fatigue and the increased complexity of questions used on a proficiency exam, a reasonable rate is 80 questions per hour. The tests currently administered through Austin Peay State University require 100 minutes divided into two 50 minute sessions with a 10 minute rest period. Our experience has suggested that this structure is adequate for examinations of 120 to 140 items or 70 items with 4 essays.

Scoring and Statistical Treatment of Outcome

Computer scoring of the test will permit a variety of statistical techniques to be employed. A routine item analysis should always be conducted and should provide information on each item, including the number of correct responses, the number left blank (particularly if a penalty is used for incorrect responses), the number of incorrect responses, a discrimination index and a difficulty index. A typical discrimination index is the point biserial correlation coefficient (Edwards, 1984) constructed between overall score for a given area and the response for the item within the area. A statistically significant positive index (such as in numbers 1, 2, and 3 below) indicates an item which was more or less consistently answered correctly by those examinees who performed well in the area and incorrectly by those who did not. The difficulty index is simply the ratio of correct to total responses (ie., the percent of correct responses). Where normative data are to be collected across several administrations, the data from the individual item analyses should also be retained. This permits the item analyses to be extended to include each additional administration without discarding the item histories.

ITEM	RESPONSE			DISCRIMINATION	DIFFICULTY
#	CORRECT	SKIP	INCORRECT	INDEX	INDEX
1	83	2	37	.4960 (p < .001)	.6803
2	56	10	56	.3788 (p < .001)	.4590
3	199	11	26	.3355 (p < .001)	.8432
4	11	9	102	.0659	.0901
5	116	3	3	.2496	.9508

On the basis of the item analysis, items can be classified according to the following scheme.

Bad - an item falls within this class if it has a difficulty index of .25 or less and the item does not have a significant discrimination index, or it has a difficulty index of .10 or less regardless of the discrimination index of the item, or the item has a significant negative discrimination index regardless of the difficulty level (see item #4 above).

Easy - an item falls within this class if it has a difficulty index greater than .90 (see item #5 above).

Normal - an item falls within this class if it does not fall within any other specified class.

Good - an item falls within this class if it has a difficulty index of .50 or greater and it has a significant positive discrimination index (see items #1 and #3 above).

Hard - an item falls within this class if it has a significant positive discrimination index and a difficulty index between .10 and .49 (see item #2 above).

Items which fall into the "bad" classification should not be included in the construction of the scores. Items which fall into this classification on two consecutive administrations of the test should be replaced at the earliest possible time.

Following the item analysis, scores can be calculated by simply awarding points for all of the valid items answered correctly and dividing the total by the number of valid items in a given area. An alternative approach is to weight the points awarded for correct answers depending upon the difficulty level of the item. The following weighting scheme is used for the Austin Peay tests.

Bad = 0.0
Easy = 0.5
Normal = 1.0
Good = 1.5
Hard = 2.0

The area scores are then calculating by dividing the total weighted points earned by the total weighted points available in the area. This technique also provides a means by which to compensate for the different difficulty levels encountered when using multiple forms of an examination.

In order to make norm based comparisons of test outcomes, once the raw scores have been calculated they should be converted to $\underline{Z}$ scores using the mean and standard deviation of the overall normative sample. To facilitate interpretation, these $\underline{Z}$ scores can be converted to an arbitrarily determined score base. The Austin Peay score base is 500 with a standard deviation of 100. The $\underline{Z}$ is multiplied by 100 and added to the score base to produce a scoring system with a mean of 500 and a standard deviation of 100. This greatly eases intepretation of the results of the test. Once the individual area and overall scores have been calculated, departmental averages can be constructed.

Other Statistical Treatments of the Results

Numerous other statistical treatments can be applied to the test results in order to provide additional feedback to academic departments concerning their students' performance. For example, product-moment correlation coefficients can be calculated between the various areas tested so as to determine the relative contributions of each area to the other areas and to the overall score. If the students are asked to complete a check list of the areas in which they have taken courses, point-biserial correlations can be calculated between whether or not a student has taken a course in a given area and performance on that portion of the test. Other treatments can be applied depending upon the needs of the individual department. A sample report has been included at the end of this paper.

Summary

Although the role of legislatively mandated outcome assessment in the educational process is anything but clear at this time, the requirements continue to be applied in many states. Whatever solution an institution elects to use to meet the requirements will require time, effort and funding. Experience with the design and administration of "in-house" senior proficiency tests in several different disciplines has shown, however, that the tests can be quite valid and can be used to provide considerably more than the minimum mandated "overall performance score." Furthermore, the enhanced feedback can be obtained from a test costing one-half to one-third as much as a commercially available instrument. These examinations can be constructed by an individual department or as part of a consortium of institutions and easily can be evaluated in terms of statistical validity and reliability.

```
INSTITUTION 999:  Midlothian University
Prepared:  2-JUN-86            Test Year:  1986        Field:  Psychology
---------------------------------------------------------------------------
SAMPLE SIZE FOR THIS INSTITUTION >   35
---------------------------------------------------------------------------
     AREA TESTED                         MEAN        STANDARD     % ILE     CASES IN
                                    STANDARD SCORE   DEVIATION             NORM GROUP

( 1) Experimental Design                 492.0         113.4       46.8        236
( 2) Statistics                          472.8          77.0       39.3        236
( 3) Physiological                       493.9          98.5       47.6        236
( 4) Animal Learning/Motivation          506.6         106.4       52.6        236
( 5) Social                              499.7          90.8       49.9        236
( 6) Human Learning/Cognition            496.2          72.9       48.5        236
( 7) Personality                         489.3          98.8       45.8        236
( 8) Abnormal                            470.8          86.7       38.5        236
( 9) Developmental                       514.7          93.8       55.9        194
(10) History & Systems                   467.2          98.7       37.1        114
(11) OVERALL SCORE                       489.0          83.4       45.6        236
---------------------------------------------------------------------------
INTERCORRELATIONS OF AREA SCORES AND AREA SCORES WITH OVERALL SCORE

        ( 2)    ( 3)    ( 4)    ( 5)    ( 6)    ( 7)    ( 8)    ( 9)    (10)    (11)
( 1)  0.0224  0.0931  0.2494  0.1113  0.0119  0.2387  0.2158  0.3348  0.3789  0.5101
( 2)         -0.0573  0.3016  0.1629 -0.1130  0.1147  0.3092  0.1109  0.2666  0.3878
( 3)                  0.0823  0.3287  0.2861  0.2466  0.1581  0.0938  0.2802  0.5007
( 4)                          0.0884  0.0228  0.2041  0.1931  0.1780  0.1836  0.4759
( 5)                                  0.1092  0.4884  0.3202  0.2705  0.2247  0.6300
( 6)                                          0.3367 -0.0221 -0.1559 -0.2472  0.1990
( 7)                                                  0.3105 -0.0719  0.1220  0.5856
( 8)                                                          0.4758  0.5476  0.6940
( 9)                                                                  0.3759  0.5164
(10)                                                                          0.6242

---------------------------------------------------------------------------
POINT BISERIAL CORRELATION COEFFICIENTS AND ASSOCIATED t VALUES FOR
COURSES TAKEN/NOT TAKEN WITH CONTENT AREA SCORES

     AREA                                DF                R          t

( 1) Experimental Design                 33             0.0809     0.4663
( 2) Statistics                          33            -0.0758    -0.4367
( 3) Physiological                       33             0.5351     3.6383     p<.05
( 4) Animal Learning/Motivation          33            -0.0477    -0.2742
( 5) Social                              33             0.1043     0.6024
( 6) Human Learning/Cognition            33             0.2129     1.2518
( 7) Personality                         33             0.3095     1.8700
( 8) Abnormal                            33            -0.2215    -1.3048
( 9) Developmental                       33             0.2497     1.4816
(10) History & Systems                   33            -0.0508    -0.2921
```

References

Bloom, B., Madaus, G., and Hastings, J. (1981). Evaluation to improve learning. New York: McGraw-Hill Book Company.

Edwards, A. L. (1984). An introduction to linear regression and correlation. New York: W. H. Freeman and Company.

Dr. Anthony J. Golden is with the Psychology Department, Austin Peay State University, Clarksville, Tennessee 37044 USA

ASSESSMENT IN UNDERGRADUATE EDUCATION AND SOME IMPLICATIONS FOR MINORITIES

Freeman A. Hrabowski

In a recent article about objective tests and instruction in composition, Charles Suhor, deputy director of the National Council of Teachers of English and director of the ERIC Clearinghouse on Reading and Communication Skills, cited an interesting allegory created by Linda Darling-Hammond. We are suddenly transported to Wonderland, where a prestigious national commission has just called for major medical reforms.

> In response, a major hospital decided to institute performance measures of patient outcomes and to tie decisions on patient dismissal as well as doctors' salaries to those measures. The most widely used instrument for assessing health in Wonderland was a simple tool that produced a single score with proved reliability. That instrument, called a thermometer, had the added advantage of being easy to administer and record....
>
> When the doctors discovered that their competency would be judged by how many of their patients had temperatures as measured by the thermometer as normal or below, some complained that it was not a comprehensive measure of health. Their complaints were dismissed as defensive and self-serving. The administrators, to insure that their efforts would not be subverted by recalcitrant doctors, then specified that subjective assessments of patient well-being would not be used in making decisions. Furthermore, any medicines or treatment tools not known to directly influence thermometer scores would no longer be purchased.
>
> After a year of operating under this new system, more patients were dismissed from the hospital with temperatures at or below normal.... Some years later, during the centennial Wonderland census, the census takers discovered that the population had declined dramatically and that mortality rates had increased. As people in Wonderland were wont to do, they shook their heads and sighed, 'Curiouser and curiouser" (Suhor, p. 63).

Recent reports by national and regional groups examining the quality of American education -- among them the Southern Regional Education Board's Access to Quality Undergraduate Education and the National Institute of Education's Involvement in Learning -- have identified deficiencies in the undergraduate curriculum and have recommended that colleges adopt assessment or competency testing programs in order to address these deficiencies. At the same time, state legislatures, industry, educational coordinating boards, and accreditation agencies have been insisting that institutional accountability be defined by outcomes that can be clearly measured. Because of the value placed on accountability and the increasing significance of the competency testing movement in higher education, it is important that we examine the nature of this movement and assess its effect on undergraduate education.

Since the mid-1970's, forty states have mandated competency tests in their public schools: approximately half of these states have made passing the tests a requirement for high school graduation. This new emphasis on testing has come from state legislatures, state departments of education, and state boards of education, all of which share the responsibility for ensuring that schools are of high quality. Since these people also share some of the responsibility for public higher education, and since one report after another has concluded that American college curricula have decayed, it is not surprising that states have begun to ask questions about the quality of public higher education and have begun to require evidence of quality. And, it is standardized tests that provide the basis for that evidence.

Though competency testing programs differ, they share several characteristics. All require students to attain a cut-off score that represents the standard at and above which students are considered competent. And all have possible significant consequences for students, including remediation, progression, graduation, job placement, and career opportunities.

At present, only a handful of colleges and universities have implemented comprehensive competency testing programs, generally as a result of state initiatives (Ewell, p. 32). While assessment is certainly not new in higher education, the objectives of the current movement are distinct from traditional assessment objectives. First, traditional testing is used primarily to determine an individual student's mastery of specific course content; test scores of aggregates are rarely considered important. In contrast, the current competency testing movement focuses on the evaluation of institutions and programs through aggregate student testing. Second, in traditional testing the decision to advance a student to the next level is determined internally by the professionaljudgment of the faculty, but in the current competency test movement that decision is often made externally by state boards which adopt testing requirements and set cut-off scores. For instance, "rising junior" testing programs in Georgia and Florida require all sophomores in public colleges and universities to pass academic skills tests before they can advance to the next level. In the Florida program students must pass all four portions of the State's College-Level Academic Skills Test before they can continue as juniors or earn an associate of arts degree. Students may repeat the test until they pass it, and those who pass three of the four portions may take up to 36 credits in upper division courses before they are required to pass the fourth part. The cut-off scores for passing the test will be raised in 1986 and 1989. When the test was initiated in 1984-85, students were required to answer correctly only about half of the test questions; by 1989 they will have to answer 70% correctly. The initial cut-off score was deliberately set low in order to provide time for reforms in elementary and secondary education. Interestingly, if the 1989 cut-off score had been used in 1984, only 45% of Florida's students would have passed instead of the 84% who actually did (Evangelauf, p. 10).

Because of the public's lack of confidence in education, the public itself views testing requirements mandated by external agencies as a welcome approach to restoring high standards in education. We can expect that educators will eventually have to accept decreased reliance on professional judgement, increased dependence on standardized tests, and increased regulation of education at all levels by the states.

On the one hand, academicians generally agree that colleges and universities have not developed measures for evaluating their own effectiveness, that there is a need for substantive improvement in the quality of undergraduate instruction, and that assessing the outcomes of the college experience can help institutions to improve. "Without some accurate sense of the progress with which students are establishing skills and mastering the capacities defined by the minimum required curriculum, a faculty can only guess at how well it is doing its job" (AAC, p. 33). On the other hand, many academicians are understandably concerned that standardized testing programs mandated or imposed on institutions externally should take into account the vast differences among colleges and universities; and, correlatively, that assessment of programs should take into account students' pre-college academic background since students enter college with varying levels of ability and skills. The "value-added" approach to assessment, for instance, requires pre- and post-testing of students with the same types of instruments. Determining the effectiveness of a curriculum, then, would mean not only looking at what students know when they graduate but also considering what they knew when the entered.

Given this simple observation, how do colleges, in fact, measure gains? Most competency testing programs focus on cognitive and skills development outcomes. With regard to cognitive development, the distinction is made between general knowledge and the knowledge associated with specific disciplines. Many colleges and universities administer the SAT or ACT entrance examinations and the Graduate Record Examination to determine gains in general knowledge. A number of other institutions use the ACT College Outcome Measures Project (COMP) to assess students' knowledge of general education. The GRE Field Examinations are frequently used to measure students' knowledge in specific disciplines, and CLEP examinations are used to assess course-level knowledge.

These tests purport to measure what students actually know. Yet, in recent years the public has been concerned not only about what students know but also about what students can do and how well they can do it. Consequently, competency testing programs also focus on writing, speaking, mathematics, and critical thinking. The more specialized skills associated with particular occupations are frequently tested using standards of performance and instruments established through certification processes.

Now that we have examined the pressure on colleges to use competency testing and discussed various kinds of standardized tests that resulted from these pressures, we need to focus on ways in which faculty at various colleges and universities have reacted. Most important, faculty have been using data gathered from standardized tests for seniors, rising junior examinations, and value-added tests to review and improve curricula. Many faculty in assessment programs report devoting more time to examining curricula, discussing the content of courses, and reflecting on what they expect students to accomplish than was the case before these programs were developed. Several examples illustrate this process. First, in response to a state mandate for assessment, the University of Tennessee, Knoxville, used the results of a senior test to change freshman course requirements (Ewell, p. 35). In the value-added assessment program at Northeast Missouri State University, when faculty detected only limited growth in mathematics skills amoung students during their first two years, they developed programs of mathematics across the curriculum which rcsulted in marked improvements within three years. When psychology faculty found that students with high grade point averages scored poorly on the GRE psychology exam, they realized, after reviewing student test scores and transcripts, that students were completing major requirements without taking certain critical basic courses. After the department responded by changing requirements, GRE scores increased (Jaschik, p. 22).

The experiences of these and other institutions suggest that in areas involving specified basic skills and well-defined bodies of knowledge, student achievement can be improved through the use of standardized tests as a basis for curriculum revision.

However, faculty use of standardized tests that measure general knowledge has not produced such clear results. For example, the results of such examinations as the ACT-COMP, which tests students' knowledge of general education and the liberal arts, are more difficult to analyze and use for curriculum revision. Moreover, educators would agree that to evaluate the quality of a college education, or even of the first two years, on the basis of success in teaching basic skills would be to confuse the work of the college with that of the public schools. Tests of basic skills provide no information about the student's knowledge of literature or history, or about his character or his understanding of others. Unfortunately, beyond the level of basic skills there is considerable disagreement about the goals and purposes of general education as seen in the marked diversity among American colleges and universities.

In addition, the content of tests such as the GRE sometimes does not closely match the focus of a major program. For example, several departments of English at universities in the midwest have not been satisfied with the GRE field examinations because these tests do not reflect the departments' strengths in American literature: consequently, it would be difficult to determine the relationship between students' test performance and their achievement in the major department.

Even with the demonstrated effectiveness of some testing programs, many faculty continue to have their doubts about the long-term effects of standardized tests. Most English professors who read the allegory at the beginning of this paper would associate the thermometer with objective, multiple choice tests of writing ability and associate the doctors with instructors of writing who, understandably, argue that required objective tests are not adequate and sufficient as assessment implements. Like the doctors in the allegory. a number of faculty believe that the short-term gains in scores will be overshadowed by the potentially disastrous long.term effects of objective tests on students' writing skills. Nonetheless, there does seem to be a growing trend to use writing samples as a part of standardized tests, and according to many instructors of composition the use of these samples is likely to ensure a more appropriate diagnosis of students' writing skills (Suhor, p. 64).

In general, however, critics have charged that standardized tests

> ...measure only a narrow spectrum of abilities, that tests by their very nature discourage creative and imaginative thinking that the results of the tests have far too significant an effect on the life chances of young people, that the emphasis in a multiple-choice test is wrongly on "the right answer" and on simplicity instead of thoughtful judgements, that the tests favor the advantaged over the disadvantaged while claiming to be neutral, and that the tests are inherently biased against those who are unfamiliar with the language and concepts of the majority culture (Ravitch, p. 59).

The increasing reliance on competency tests in Maryland higher education is readily apparent from the December 1985 draft of "The Maryland Statewide Plan for Postsecondary Education." According to this document, the state will work with Maryland colleges and universities to develop a testing program for all students completing their sophomore year (p. 23). The new emphasis on testing in the state is also clear from the recent decision of the Maryland State Department of Education to adopt the National Teacher Examinations as a requirement for certification, effective Spring 1987. These planned testing programs simply reflect the national trend. As of August 1985, 28 other states had adopted testing requirements for teacher certification, and 31 states had testing requirements for admission to teacher education programs.

For most students, these competency tests present no special problems since only minimum level skills are tested. However, the teacher testing issue is particularly heated because of the poor performance of minorities on these tests. For example, in 1982, 85% of Louisiana's black applicants failed the NTE compared to 22% of whites. In California in 1983, 62% of all students who first took the certification examination passed, as compared to only 29% of minorities (Dilworth, pp. 37-39). In Florida, reports for 1982-83 on the Florida Teacher Certification Examination show a 90-92% pass rate for white, 35-37% for black, and 51-57 for Hispanic candidates (Garcia, p. 5). In Georgia in 1983, 34.4% of the black candidates passed compared to 87% of the white candidates. In fact, devastating statistics on the disproportionately large number of minority applicants who fail teacher competency tests can be found in all states that have implemented such requirements. More devastating still is the fact that some states have begun using test results to make decisions about teacher training programs. For example, "...in July 1983, Florida decertified 38 of its 436 teacher training programs because of high student failure rates" (Dilworth, p. 38).

In anticipation of Maryland's decision to adopt a testing requirement for teachers, Coppin State College, a predominantly black college in Baltimore, began working with the Southern Regional Education Board and the Educational Testing Service to develop a project titled "Improving the 'Pass Rate' of Minority Students on the National Teacher Examinations and Other Standardized Tests." This project is underwritten by the Fund for the Improvement of Post Secondary Education for the period 1984-1987. It focuses on faculty development, curriculum development, student assessment, and student development. Approximately 25 faculty (representing arts and sciences departments and education departments) are participating in on-going workshops conducted by consultants from the Educational Testing Service and experts from other institutions. In addition to developing their skills in item analysis, test construction, and test-item writing, these faculty have been reviewing selected sections of sample tests in order to determine the extent to which existing courses in our curriculum can equip students with those skills needed to perform successfully on the test.

Let me emphasize four points: 1) we are *not* merely teaching the tests; 2) we are *not* limiting our curriculum to the content of the tests; 3) we *are* integrating the teaching of test-taking skills into regular classroom instruction; and 4) we *are* gearing instruction and testing to foster the higher order skills of writing and thinking.

Coppin faculty are also participating in reading and writing workshops designed to teach them how to assist students with reading and writing problems within specific disciplines. Though it will be years before we know the actual impact of this project on the test performance of students, the short-term benefits are already apparent. The workshops have provided excellent opportunities for faculty to talk about our curriculum, its structure, content, and general objectives, and faculty are talking with colleagues from other departments about ways of working together to reinforce concepts. For example, faculty who teach math and logic courses are collaborating on the teaching of transitive inference, deductive reasoning in a natural language, and flow charting, since the math section of the NTE concentrates on general reasoning as well as numerical computations. Finally, at least now faculty know what is expected of students on the NTE, and many are constructing tests which require of students not only familiarity with the format of standardized test questions, but also strong reading, writing, and critical thinking skills.

Let me close by quoting from the NIE report Involvement in Learning Realizing the Potential of American Higher Education.

> ...we are interested in assessment as a means to provide information about the teaching and learning process and as feedback to help improve the effectiveness with which students, faculty, and the institution carry out their work. We thus believe assessment to be an organic part of learning.
>
> The use of tests to sort and screen is legitimate for professional certification and licensing -- and, indeed, for any operation where selection is necessary. But assessment has even greater potential as a tool for clarifying expectations and for increasing student involvement when it is used to measure improvements in performance.
>
> ...the values of institutions are often revealed by information they gather and use. If a college regularly seeks longitudinal data on how much students are actually learning and developing, then that institution explicitly places a high value on its responsibility to change and educate its students (p. 53).

BIBLIOGRAPHY

Association of American Colleges (AAC). Integrity in the College curriculum A Report to the Academic community. Washington, D.C., 1985.

Dilworth, Mary E. "State Status Report." In Teachers' Totter, A Report on Teacher Certification Issues. Washington, D.C. Institute for the Study of Educational Policy, 1984.

Evangelauf, Jean. "Sophomores in Florida Public Colleges Must Pass a New Test of Academic Skills." The Chronicle of Higher Education, 3 July 1985: 10.

Ewell, Peter. "Assessment -- What's It all About?" Change, November/December 1985: 32-36.

Garcia, Peter. "A Study on Teacher Competency Testing and Test Validity With Implications for Minorities and Results and Implications of the Use of the Pre-Professional Skills Test (PPST) as a Screening Device for Entrance into Teacher Education Programs in Texas." Report to National Institute of Education. Washington, D.C., 1985.

Jaschik, Scott. "As States Weigh 'Value-Added' Assessment, Northeast Missouri Offers Model." The Chronicle of Higher Education, 2 October 1985: 22.

National Institute of Education. Involvement in Learning: Realizing the Potential of American Higher Education, Washington, D.C.: 1984.

Ravitch, Diane. "Value of Standardized Tests in Indicating How Well Students Are Learning." In the Uses and Misuses of Tests. Ed. Charles W. Davis. San Francisco, Jossey-Bass Inc., 1984.

Salganik, Laura Hersh. "Why Testing Reforms Are So Popular and How They Are Changing Education." Phi Delta Kappa, 66 (May 1985), 608.

Suhor, Charles. "Objective Tests and Writing Samples: How Do They Affect Instruction in Composition?" Phi Delta Kappa, 66 (May 1985): 635.

Freeman A. Hrabowski is Vice President for Academic Affairs at Coppin State College, 2500 W. North Avenue, Baltimore, Maryland 21216

Miami-Dade Community College's Use of the College Level Academic Skills Test (CLAST)

J. Terence Kelly and John Losak

Purpose and Status of This Plan

Test Administration Plan sets forth the terms and tions which govern the administration of the College-Academic Skills Test (CLAST) in community colleges and state universities as well as in any other institutions which may be authorized to participate in the testing program. The Test Administration Plan consists of this document and the 1985-86 Test Administration Manual which is made a part hereof but which is published as a separate document.

The Test Administration Plan, including the Test Administration Manual, is intended to ensure that all students who take CLAST have equal opportunity to demonstrate the level of their achievement of the skills which are measured by the test.

This plan has been developed and issued as an exercise of the responsibility for statewide administration of CLAST which is vested in the Commissioner of Education by Section 229.551 (3) (k), Florida Statutes, 1983. Further, it is noted that the test, test scoring criteria, and testing procedures are exempt from rule-making requirements by provisions of Section 120.52 (15) (e), Florida Statutes, 1983.

Effective Date

The Test Administration Plan for 1985-86 becomes effective on the date of issuance by the Commissioner of Education, and it replaces the Plan for the Administration of the CLAST, issued August 15, 1981 and all supplements thereto.

Overview

The College-Level Academic Skills Test (CLAST) is a part of an overall effort in Florida which is designed to assure that students completing any level of education have acquired the skills which are expected of them. It is the particular function of CLAST to measure the achievement of the communication and computation skills which are expected of all students by the time they complete their sophomore year in college.

The skills which are measured by CLAST were identified by the faculties of community colleges and state universities through the College-Level Academic Skills Project; the skills have been adopted by the State Board of Education and are listed in Rule 6A-10.31, FAC. Provisions for keeping the skills list current and for maintaining active participation of faculty members in the implementation of the testing program are set forth in Appendix C.

Since August 1, 1984 scores on CLAST which satisfy the standards of the State Board of Education Rule 6A-10.312, FAC, have been required for the award of an Associate of Arts degree from a community college or state university and for admission to upper division status in a state university in Florida.

Inasmuch as participation in CLAST is required under provision of State Board of Education Rule 6A-7.17, FAC, as a condition of eligibility for State of Florida financial aid awards, CLAST is made available to students at private colleges and universities, at the initiative of the respective institutions.

Statutes and rules which are applicable to the CLAST requirement are contained in Appendix B.

Purpose of CLAST

The College-Level Academic Skills Test has been designed as an achievement test. It is intended to measure the level of achievement of the communication and computation skills which are expected of all students by the time they complete their sophomore year in college, i.e., those which are listed in State Board of Education Rule 6A-10.31, FAC.

Each of the four subtests of CLAST--computation, reading, writing, and essay--has been designed to yield a single score which is a valid and reliable estimate of the achievement by individual examinees of the group of skills measured in that subtest. For individual students the objective subtests also provide additional information by broad skill area which is useful in identifying areas of possible strength or weakness. While the test is not designed to yield information, skill by skill, needed for full diagnosis of the problems of individual students, institutions can use the results aggregated across administrations by broad skill area with confidence in determining the extent to which they have been successful in teaching their students as a group the skills measured in CLAST.

While there is support for the presumption that CLAST scores relate positively to other measures of academic performance, both prior to and after the test has been taken, CLAST does not purport to be a predictor of subsequent performance of examinees in upper division programs.

Description of the Test

The test consists of four subtests--reading, writing, computation, and essay. The reading subtest measures eleven skills in the areas of literal comprehension and critical comprehension. The objective writing subtest measures twelve of the writing skills in the broad areas of word choice, sentence structure, and grammar, spelling, and punctuation. The essay affords students opportunity to demonstrate all twenty-four of the writing skills by writing an essay on one of the two topics presented. The computation subtest measures fifty-five skills in the areas of arithmetic, algebra, geometry and measurement, logical reasoning, computer technology, and statistics, including probability.

The number of items by subtest and by broad skill areas within each subtest is shown in Appendix D. Equivalent forms of the test which meet these specifications are developed and equated for each administration.

The test consists of two books.-one containing computation items and one containing reading and writing items, as well as instructions for the essay. In order to increase test security, multiple forms of each test, each using a different arrangement of items, are printed for each administration. Braille, cassette, and large-print versions are available.

At the close of each academic year a technical report is issued to provide detailed technical information concerning the development of CLAST and of its performance to date.

Testing Schedule

The test is administered in one morning session, which includes a short break. Actual testing time is three and one-half hours, plus time required for checking in examinees, coding identifying information, distributing and collecting materials, and reading directions for each subtest. The order in which subtests are administered and the time allowed for completion of each is shown below:

Essay	50 minutes
Reading and Writing	70 minutes
Computation	90 minutes

Modifications in the testing schedule, administration procedures, and/or format for handicapped examinees are detailed in the Test Administration Manual.

Responsibility for the Administration of CLAST

The College-Level Academic Skills Test is a standardized test which must be administered across the state in a uniform, secure, and professional manner. That requirement is dictated by the fact that the consequences of the test to students are such that all examinees must be afforded equal opportunity to demonstrate the level of their attainment of the skills measured by the test. In order to ensure that all examinees take CLAST under the same conditions, detailed assignments of responsibility are made in the Plan to each of the three levels involved in the administration of CLAST. the CLASP Office, the Statewide Test Administrator, and the Institutional Test Administrator.

1. The CLASP Office of the DOE

By law the Commissioner of Education must maintain statewide responsiblity for the administration of the College-Level Academic

Skills Test; and the functions which cannot be assigned to the Statewide Test Administrator are exercised by the CLASP Office of the Department of Education. The responsibilities assigned to the CLASP Office are as follows:

a. Serving as the public spokesperson for the testing program
b. Developing the Test Administration Plan
c. Approving Institutional Test Administrators
d. Negotiating contract(s) to effect the assignment of responsibility to the Statewide Test Administrator
e. Monitoring and verifying performance of the State-wide Test Administrator under terms of the assignment
f. Monitoring the review of skills by the College-Level Academic Skills Project and recommending any proposed changes in the list of skills to the Commissioner of Education
g. Approving specifications for the development of test items
h. Negotiating and monitoring contracts for the development of test items
i. Reviewing and approving test blueprints
j. Reviewing and approving draft and camera-ready copy for all test materials prior to printing
k. Verifying the accuracy of computer programs and score reports
l. Validating test score keys
m. Providing and orienting personnel authorized to observe the administration of the test
n. Investigating alleged irregularities in the implementation of the Test Administration Plan or in the administration of the test, including any alleged breaches of security
o. Reviewing and processing complaints from individuals or institutions
p. Approving the Test Administration Manual
q. Providing content for student information brochures and posters
r. Releasing statewide test results and other information relating to the testing program
s. Approving the Technical Report
t. Reviewing and monitoring the impact of CLAST standards

2. The Statewide Test Administrator (STA)

Under the law the Commissioner of Education is authorized to assign administrative responsibility for the College-Level Academic Skills Test to any public community college or state university. Such assignment has been made to the Office of Instructional Resources of the University of Florida and the assignment has been effected through contracts which assign responsibilities as follows:

a. Developing and maintaining the item bank for CLAST
b. Developing blueprints for forms of the test
c. Designing and formatting test books, the answer folder administration manual, score reports, score interpretation

guide, student registration forms, and student information posters and brochures

d. Printing all testing materials
e. Disseminating test materials and providing for their return
f. Training Institutional Test Administrators
g. Directing the administration of CLAST
h. Scoring tests, including the essay
i. Providing score reports to individual examinees
j. Providing reports of test results to institutions and to the Department of Education
k. Analyzing test results and recommending changes in the test and/or test procedures
l. Preparing a technical report
m. Storing test materials as required by the record retention schedule for CLAST materials

3. Institutional Test Administrator (ITA).

One individual in each participating institution is assigned responsibility for the coordination of all activities in that institution relating to the CLAST testing program. The appointment is initiated by the president of the institution with the nomination of an individual who is a qualified test administrator, experienced in the administration of secure, standardized examinations. The nomination will include a resume to document the experience and training of the nominee.

While an individual who serves as ITA for CLAST may have additional institutional duties, the written job description in the institution must assign responsibility for all of the functions listed below to that individual. The individual nominated should be so situated in the institutional structure as to facilitate the effective discharge of all of those responsibilities.

When approved by the Department of Education, the ITA becomes the principal point of contact between the Department of Education on all matters relating to CLAST and will be held responsible by the Department of Education for the discharge of the specific responsibilities as follows:

a. Notifying students of the examination and of the requirements for CLAST scores
b. Receiving applications from students to take CLAST
c. Determining the eligibility of applicants to take CLAST
d. Giving room assignments and admission tickets to students who have been found to be eligible to take the test
e. Identifying and scheduling test rooms according to the requirements set forth in the Test Administration Manual
f. Appointing room supervisors and proctors as stipulated in the Test Administration Manual
g. Training room supervisors and proctors in the use of materials provided by the Statewide Test Administrator

h. Determining what special testing conditions are to be provided for examinees who have handicaps
i. Notifying the Statewide Test Administrator, upon the close of registration for the test, of the exact number of students registered and ordering test materials in the manner detailed in the Test Administration Manual
j. Observing all requirements for test security which are set forth in the Test Administration Manual
k. Insuring that the administration of CLAST in each test room is governed in detail by the Test Administration Manual
l. Determining the eligibility of students to participate in the make-up administration
m. Certifying to the CLASP Office of the DOE any students who are eligible to participate in the central make-up administration of CLAST
n. Making arrangements for eligible students who request to take CLAST in an institution other than their own
o. Certifying to the CLASP Office of the DOE the eligibility of students who request out-of-state administrations of CLAST
p. Receiving and disseminating summary reports of test results in their respective institutions

J. Terence Kelly is Vice President, Miami-Dade Community College North, 11380 N. W. 27th Avenue, Miami, Florida 33167

John Losak is Dean of Institutional Research, Miami-Dade Community College, 11011 S. W. 104th Street, Miami, Florida 33176

EVIDENCE OF QUALITY ASSESSMENT AND CONTROL IN STATE BUDGET DOCUMENTS

by Jane Adams Lamb

The funding of public universities and colleges in America has historically been a matter left to the states, an issue resolved in a climate of respect for higher education and of compromise based on fifty different political realities. Examination of the complex budget documents for higher education in each state show that no two states justify the dollars spent for higher education in exactly the same way. Yet within those differences, one can identify similarities and trends that reveal not only expenditure patterns, but philosophical patterns affecting public institutions today. If the predicted influences on higher education in the eighties--enrollment decline, increasing government intervention, declining state resources, and increasing emphasis on quality and accountability [6, 19, 18] are occurring, they can be expected to be evident in the budget documents.

The primary tool utilized for the allocation of state funds for public higher education is the formula, a descriptive compilation of apparently objective mathematical relationships or "formulas" that relate historic or projected cost to work load and that yield an estimated future funding need. [5,11,12] Yet, none of the various national statistical data reports on higher education regularly compiles data on state budget formulas in sufficient detail that the researcher can identify particular differences between states. Due undoubtedly to the complexity and diversity of these patterns, as well as the frequency of change, this lack prevents examination of a valuable source of information on public higher education.

This study examined state budget request documents in order to identify changes that were occurring in budgeting methods. Differentiation between "formula" and "non-formula" states was often difficult. As Meisinger suggests, "the definition of a budget formula is a debatable issue among both practitioners and researchers alike." [15, p. 23] States were counted as "formula states" if the responses specifically described and/or labeled an objective mathematical budget process with institutional measures "a formula." Data was collected on the structure of these budget formulas used in 1985 to generate budget requests or allocations to public higher education by conducting a mail survey of the higher education coordinating bodies in forty-eight states. In Wyoming and Vermont the request was made directly to the university system. Copies of the budget formulas or guidelines utilized in each state were solicited. In addition to the budget documents requested, four questions were asked: (1) What percentage of the formula request was actually funded in the last funding period? What was the average percentage funded over the last five years? (2) Do unexpended funds revert to the state treasury at the end of the fiscal period? (3) What higher education items are funded outside the formula? and (4) What system for rewarding quality or measuring performance for budget purposes is not clearly described in the formula?

Responses were received from forty-two states. Of these forty-two, twenty-seven sent descriptions of a budget or allocation formula for four-year colleges and universities. Three additional ones sent data on formulas utilized for their community colleges only. Subsequent phone calls confirmed that three of the eight nonrespondent states used budget formulas, but sufficient information was not available so that an adequate description of these three formulas could be made. Thirty-one states, a combination of formula and nonformula ones, sent some level of response to the four questions.

A comparative analysis of the structure of existing state budget or allocation formulas was then made utilizing the content of the state documents and letters. Particular aspects of the formula process or structure, especially those that had been identified in the current literature as changing, were examined. The structure and complexity of present formulas was then compared to the formulas described by Dr. Frances Gross in his similar surveys in 1973 and 1979. [11, 12] Factors examined in this comparative analysis were the number of states using formulas, the base factors utilized, the computational methods, the level of differentiation within formulas, the complexity of the formulas, and the geographic distribution of states using formulas.

Description of Formula Budgeting

Thirty states were using formulas in the preparation of budget requests for state funds for four-year public colleges and universities and/or in the distribution of allocated state funds within the higher education system. There was no automatic linkage between the budget preparation criteria and process and the allocation criteria and process. Four states described a formula used for allocation distribution only. Two states described two separate formulas--one for the budget request and one for the allocation distribution. The remaining twenty-four states reported on formulas used for the budget request process only. The participating parties in both of these processes varied from state to state. The formulas were designed and sanctioned by the university system itself, the higher education coordinating body of the state, the governor, the legislature, or any combination of the above.

Evidence of interest and examination of budget formulas was apparent by the number of state studies with resulting changes that had occurred since 1980 (fourteen) and the number of studies underway in 1985 (six). Yet, with all the changes occurring, formulas continued to be structured primarily by the functional classification areas that define the Educational and General portion of the budget (Instruction, Research, Public Service, Academic Support, Libraries, Student Services, Institutional Support, and Operation and Maintenance of Plant). Instruction and Operation and Maintenance of plant were the two categories most often included within formulas.

Base factors that drove the formula (in order of frequency of use) were portions of current or past budgets, student credit hour production, FTE enrollment, square feet/acreage, head count, FTE faculty/staff positions, and value of inventory. The most frequently used base factor was the budget itself, either the previous year's budget or some portion of the present proposed budget, continuing the emphasis of an incremental approach to budgeting and masking the importance of the base factor driving the other portion of the budget. If FTE enrollment, head count, and student credit hour production were considered jointly, then the number of students within the educational system became by far the greatest factor in determining the budgeted amount. Thus, states did not appear to be abandoning enrollment-driven formulas in times of stability or decrease in numbers of students.

Yet a variety of methods were being used that mitigated the power that current enrollment had on the final formula figure. Use of head count, especially in the area of Student Services, occurred in fourteen states. Buffering or decoupling was described in some portion of the formula in sixteen states, and capping of enrollment was mandated in the budget documents in only two formula states. Buffering describes the various processes that limit the responsiveness of the budget to enrollment changes by redefining the amount of change that will be taken into account. For example, multiyear averages, limits on allowable or recognized growth or decline, and corridor or threshold approaches fall in this definition. Under typical corridor or threshold approaches, a range of enrollment is established, and only changes outside this range affect funding. Decoupling involves removing the linkage to enrollment altogether in as many parts of the formula as possible. [13] Marginal costing was being used in only 4 states, whereas fixed and variable costing techniques appeared to some extent in almost every formula. Six states utilized fixed and variable costing techniques in every portion of their formula. The inclusion of summer school hours in the FTE or SCH count appeared to be increasingly popular.

The three computation methods defined by Dr. Francis Gross in 1973 [11] continued to describe the varieties of calculations within formulas. The rate per base factor unit (RBFU) describes a given rate or factor multipled by an institutional descriptor or base factor. The percentage of base factor (PBF) utilizes a given percentage of another functional classification or budgetary portion. The base factor position ratio with salary rates (BFPR/SR) establishes a ratio (as faculty/students, faculty/supporting staff, staff/square foot, faculty/credit hours) on which to calculate the needed full time equivalent positions to be multiplied by given salary rates. The base factor position ratio with salary rates (BFPR/SR) method was used most often in the area of Instruction, but the simpler rate per base factor unit (RBFU) and percentage of base factor (PBF) approaches were used more often overall. The RBFU approach was utilized particularly for calculating Operation and Maintenance of Physical Plant.

The use of data from "comparable institutions" was identified as a new item in many formulas. Although the southern states had previously used SREB standards, more formulas were citing national standards from "prestigious" institutions. The establishment of higher dollar factors, better student/faculty ratios, and more generous salaries within the formulas was attributed to and justified by data from "comparable institutions."

The amount of the formula request actually funded by the state legislature varied from 73 percent to 100 percent for formulas in 1985. If the average over the last five years was compared with that percentage, every formula state reporting had experienced an improvement in this percentage in the 1985 funding period. Both New Jersey and Washington cited the low percentage of past funding as a contributing factor in their decision to abandon the formula budgeting method.

Twelve states reported they did not have to return unexpended funds at the end of the fiscal period, but only five of these were budget formula states. California, Minnesota, Mississippi, New Mexico, and Tennessee stated that they do not have budget lapse. Georgia's 1982 Study on Public Higher Education recommended the university system be allowed to keep unexpended funds as a management incentive.

The portion of the budget included in the formula varied from state to state with an apparent movement to increase the items under the formula calculations. States like Kentucky and Oregon that recently wrote new formulas moved in the direction of more items, not less, included. The identification of "decision items" within the budget structure that highlight funding beyond continuation levels appeared frequently. Equipment replacement and upgrades were now appearing in formulas.

In comparing formulas from 1973, 1979, and 1985, several findings emerged. Fourteen states that were using formulas in the 1973 Gross study had clearly used a formula approach since that time: Alabama, Arkansas, Colorado, Florida, Georgia, Louisiana, Mississippi, Missouri, Ohio, Oklahoma, Pennsylvania, South Carolina, Tennessee, and Texas. Seven states (Maryland, Minnesota, Nevada, New York, South Dakota, Virginia, and West Virginia) were identified as using formulas in 1973 and 1985, but not in 1979. Due to limitations of the 1979 study, these states have probably also used formulas continuously since 1973, bringing the total of states using formulas continuously for at least twelve years to twenty-one. In these twelve years, four states (New Jersey, North Dakota, Washington, and Wisconsin) were identified as having discontinued the use of budget formulas, and nine (California, Connecticut, Idaho, Kansas, Kentucky, Massachusetts, Montana, New Mexico, and Oregon) appeared to have begun. Illinois, Michigan, and Nebraska were identified in 1985 as using a formula for community colleges only.

The "building blocks" of formulas--the base factors, the formula factors, and computational methods--appeared to have remained structurally very similar, but their content or identity changed. Head count was not reported in use as a base factor in any formula in 1973. By 1985 it was the measure of choice for Student Services, but student credit hour production and FTE enrollment remained the primary factor for other areas. The extent to which FTE enrollment, SCH production, and head count were utilized appears as great as the use of SCH production and FTE enrollment in 1973. In other words, the influence of enrollment on the size of the amount derived from the formula had not directly diminished in the past twelve years. What had changed was the process that was followed to arrive at the base enrollment number. Head count, summer school enrollments, buffering, and decoupling were mitigating the effect of any decreases in numbers. Square footage/acreage appeared in all three studies as the base factor of choice for Physical Plant Operation and Maintenance. No clear shifts in the use of the three computational methods can be seen.

The presence in the formula of weights to differentiate costs on the basis of academic disciplines, institutional differences, and student or class levels appeared to change little in the data of the three studies, except for two areas. In 1985 an increase in the use of weights for academic areas in the Instruction and Academic Support category and a decrease of the use of weights for institutional distinctions in Operation and Maintenance of Physical Plant occurred.

The complexity of the formulas was examined by use of this Complexity Factor:

Number of functional classifications
plus
Number of base factors above number of
functional classifications
plus
Number of weights applied to each base factor
plus
Number of formula factors above number of
base factors

Application of this Complexity Factor showed that four states had increased the complexity of their formulas from 1973 to 1985, and two had decreased the complexity of their formulas. The two who showed a decrease in complexity were Colorado and Florida, two states that had recently revamped their formulas with new emphases. Florida had capped enrollment and was emphasizing quality, and Colorado had constructed a model that was not tied directly to enrollment.

More formula states continued to come from the Southeast than from any other region in 1973, 1979, and 1985. The gain in number of states by the West from one in 1973 to five in 1985 was the only sizable change in geographic distribution.

Evidence of Quality Reward or Measurement

Twenty-one states (Alabama, Arizona, California, Colorado, Connecticut, Florida, Georgia, Idaho, Indiana, Kansas, Louisiana, Massachusetts, Missouri, Montana, New Jersey, Ohio, Tennessee, Utah, Virginia, West Virginia, and Wisconsin) reported some technique in place that was interpreted in their state to indicate special funding for quality. Other states may have these programs, but they did not, in their response, interpret them as "systems for rewarding quality or measuring performance for budget purposes." In addition to the twenty-one states reporting "in place" programs, South Carolina reported on a proposed Quality Improvement Initiative program to provide incentive funds as "seed capital" for program improvement. Of the states responding with specific programs, five were nonformula states: Arizona, Indiana, New Jersey, Utah, and Wisconsin.

Ten types of quality improvement programs were identified from the responses. They are:

1. Categorical one-time grants--grants given for a specific purpose to a specific area with no promise of continuation funding. The establishment of an endowment is an example of this type of quality incentive.

2. Percentage of budget total--addition of a certain percentage of the total operating budget to create incentive monies for quality improvement. The method by which these funds are distributed to campuses may be one that uses preset criteria or competition.

3. Designated area--funding specifically earmarked for particular academic disciplines to upgrade the quality of their offerings. An example of this type of grant is special funding to upgrade computer or engineering education in order to attract "high tech" industry to a state.

4. Competitive grants--monies awarded to campuses or disciplines competing within the system with proposals to upgrade quality.

5. Program review--required periodic review of academic programs that is in some way linked either to continuation funding or to enhancement funding. This is often viewed as a type of "quality control."

6. Student testing--an outcome measure of quality based on required testing of students on a designated instrument. Funding may be tied to the performance of students on the test or simply to the participation level of campuses in the testing program.

7. Other stated outcomes--in addition to or in place of student testing, any measure of quality that emphasizes outcomes or performance. For example, in community colleges, a measurement used was job placement.

8. Student scholarships--the state funding of merit (not need based) scholarships to attract and keep acdemically talented students in state and in public education, interpreted by some states as a quality improvement measure.

9. Centers of excellence--the creation and funding of centers specifically designed to foster excellence and focus on quality issues.

10. Staffing--the establishment and funding of chairs or reward systems designed to attract and keep quality faculty.

States reported the use of categorical one-time grants most often (ten states) as the tool for funding for quality. These grants included funds to "augment library holdings" in Alabama and to carry out "short-term focused strategies as defined by the Regents" in Georgia. Nine states were using a given percentage of the budget as "set aside" monies that could be used for quality enhancement (Alabama - two percent, Colorado two percent, Connecticut - one percent, Florida - one percent, Georgia - one percent, Missouri - variable by campus, Ohio - one percent, Tennessee - five percent, West Virginia - one percent). The designated purpose of these funds differed from state to state. For example, the objectives ranged from curriculum innovation and new program development in Florida to the establishment of Centers of Excellence in Colorado. Funds for designated areas, identified by six states, often overlapped these first two categories. Tennessee committed the largest percentage of resources and had a tiered system of various measures of quality by which funds were distributed. Competitive grants were discussed in seven states as tools to encourage excellence with additional funding. The model ranged from competition within designated topics (as curriculum or administrative efficiency) or within campuses (a set amount per campus to be awarded competitively).

Program review, student testing, and other stated outcomes were directed toward product or outcome measurement. Nine states used program review, three used student testing, and five used other outcome measures. In all of these cases no base funding was affected by specific performance measures, only enhancement funding. Program review was often in place as part of a program budgeting tool. For example, Oklahoma changed from a budget functional basis to an educational program basis for its entire budgeting process.

Four states emphasized scholarship programs (called Eminent Scholars Program in Florida) to retain the brightest students in the state in public education as a way of upgrading overall quality. Four states identified the establishment of Centers of Excellence, and five states included the establishment of special "chairs" or salary bonuses as incentives to the creation of quality.

Whether or not these funding mechanisms are considered part of the formula or not appears to be relatively unimportant in terms of their effect. Some states place them in the narrative description of their formula, and some states specifically identify them as "outside the formula." In either case, they are established funding policy included in the budget documents and apparently as objective as any part of the formula in shaping the allocation of resources.

Discussion

The use of formulas does not appear to be on the wane, as is sometimes conjectured in the light of decreasing enrollments. Rather states are searching for alternate ways of coming up with flexible funding mechanisms within the formula itself. Formula budgeting appears to be alive and well, increasing in complexity, and the favored budgeting method for the allocation of state funds to higher education. At the same time, however, states are building into the formula potential sources of accountability that may carry with them greater political control of the higher education enterprise. In defining the quality to be rewarded or in setting standards through the formula, state lawmakers are expanding their role in the decision making process.

The importance of descriptions of quality enhancement measures in budget documents described in this research lies not in their complete reflection of all the programs that are occurring around the country, but rather in the interpretation by state higher education officials that these moves reflect an emphasis on quality through the use of the budget in their particular states. The formula budgeting process has influenced and shaped higher education in unexpected and sometimes undesirable ways and will continue to do so in the future. [1] For example, formula budgeting and funding based on the number of students has encouraged growth and rewarded universities and colleges that keep students regardless of performance. Ideally, decisions about the components within the formula or the factors within any budgeting technique should be based on sound educational philosophy, not just on cost analysis and political negotiation.

If the traditional enrollment-driven formulas rewarded and encouraged access and growth, the addition of enhancement dollars appears designed to reward and encourage quality and competition. The selection of the criteria for excellence will be critical in determining the direction that higher education moves and should be decided with special concern for the mission of public colleges and universities and awareness of the incentives that are built into funding decisions. In rewarding quality, the measure is not as easy to derive as the FTE students or head count of enrollment. The primary base factors being used are still input measures (as salary, outstanding faculty, volumes in the library, the brightest students, equipment, particular programs, etc.), not output measures. Only six states described and rewarded student testing or other stated outcomes. But with increased output measures could come accountability tied to funding, a governmental intrusion not seen before in higher education.

Melvin George expressed concern about program review in terms of "unintended consequences" that influence program behavior [10, p 32]. The ability of the measuring to shape the thing being measured is very evident in formula funding. For example, the values inherent in differential weighting and selection of base factors cannot be ignored or justified entirely by cost studies. The formula is examined by state legislators and planners, not primarily as an affirmation of past expenditures, but as a signpost for future directions and needs. Selection of institutional behavior or student performance criteria to reward will provide unintended incentives for change.

An emphasis on quality enhancement appears to be being used in budget documents to "sell" increased funding for higher education to state legislatures, but the danger in this approach is that within a few years, the demand for demonstrable results for the money spent will be heard. Goals for state public higher education to be ranked in the upper-quartile range nationally for salaries or student/faculty ratios are being established in many states. With more and more states trying to reach these goals, increases in funding in every state offset each state's movement up the scale. Even student performance on standardized tests is being utilized to indicate quality improvement. But the tenuous linkage between the amount of dollars spent and the quality of the education [4] may cause problems in the years ahead for higher education leaders who promise too much for the dollars given.

In addition to identified quality enhancement dollars, the 1985 formulas appeared to be moving in certain similar directions that may indirectly reward quality. The use of data from comparable institutions to raise funding levels of formula factors, inclusion of more items under the formula, structured mechanisms within the budget formula instructions for special requests, and more factors reflective of differences in role and mission with particular emphasis on distinguishing among academic disciplines all have to potential to reward less for quantity of students and more for traditional academic programs that may enhance quality. An air of competition not seen in 1973 appeared in many budget documents, from the emphasis on merit scholarships to capture and keep the brightest students in state to prestigious faculty chairs to reward nationally-recognized faculty.

The linkage between increased enrollment and increased funding is fading, even in formula states [14], but the states have not abandoned enrollment as a driving factor in formulas. The changing definition of enrollment to reflect other work load realities (as summer school and head count) is enabling continued justification of increased budget askings even if enrollments are declining. The increasing percentages of the requests funded may testify to the effectiveness of communication about the needs of higher education, and the emphasis on quality, not quantity, has undoubtedly assisted in this communication with state legislatures. Whether this funding trend reflects or helps to shape educational issues like access (who should be given the opportunity to enter college?) cannot be determined, but the linkage is certainly present.

States appear to be changing formulas frequently, not only in response to changes in public higher education, but in order to shape public higher education in particular ways and to sell increased support to state legislators. If the decisions that are being made about the factors that drive the formula are harbingers of the direction of public higher education, then we can expect quality, accountability, and competition to be the driving forces in the years to come.

References

1. Allen, Richard H. "New Approaches to Incentive Financing." In Larry L. Leslie (ed.) *New Directions for Institutional Research: Responding to New Realities in Funding*. San Francisco: Jossey-Bass Publishers, 1984.

2. Allen, Richard H., and James R. Topping (eds.). *Cost Information and Formula Funding: New Approaches.* Boulder, Colorado: National Center ior Higher Education Management Systems, 1979.

3. Bogue, E. Grady. "State Agency Approaches to Academic Program Evaluation." In Eugene C. Craven (ed.) *New Directions for Institutional Research: Academic Program Evaluation.* San Francisco: Jossey-Bass Publishers, 1980.

4. Bowen, Howard R. *The Costs of Higher Education*. San Francisco: Jossey-Bass Publishers, 1980.

5. Brinkman, Paul. "Formula Budgeting: The Fourth Decade." In Larry L. Leslie (ed.) *New Directions for Institutional Research: Responding to New Realities in Funding.* San Francisco: Jossey-Bass Publishers, 1984.

6. Carnegie Council on Policy Studies in Higher Education. *Three Thousand Futures: The Next Twenty Years for Higher Education.* San Francisco: Jossey-Bass Publishers, 1980.

7. Carnegie Foundation for the Advancement of Teaching. *The States and Higher Education: A Proud Past and a Vital Future.* San Francisco: Jossey Bass Publishers, 1976.

8. Ewell, Peter. *The Self-regarding Institution: Information for Excellence*. Boulder, Colorado: National Center for Higher Education Management Systems, 1984.

9. Folger, John K. (ed.). *New Directions for Institutional Research: Increasing the Public Accountability of Higher Education.* San Francisco: Jossey-Bass Publishers, 1977.

10. George, Melvin D. "Assessing Program Quality." In Robert A. Wilson (ed.) *Survival in the 1980's: Quality, Mission, and Financing Options*. Tucson: Arizona University Center for the Study of Higher Education, 1983.

11. Gross, Francis McKenzie. *A Comparative Analysis of the Existing Budget Formulas Used for Justifying Budget Requests or Allocating Funds for the Operating Expenses of State-Supported Colleges and Universities*. Unpublished doctoral dissertation, University of Tennessee, Knoxville, 1973.

12. Gross, Francis McKenzie. *Formula Budgeting for Higher Education: State Practices in 1979-80*. Working Paper Series. Boulder, Colorado: National Center for Higher Education Management Systems, 1982.

13. Jones, Dennis P. *Higher Education Budgeting at the State Level: Concepts and Principles*. Denver, Colorado: National Center for Higher Education Management Systems, 1984.

14. Leslie, Larry L., and Garey Ramey. "State Appropriations and Enrollments." *Journal of Higher Education*, 57 (January/February 1986), 1-19.

15. Meisinger, Richard J., Jr. *State Budgeting for Higher Education: The Uses of Formulas*. Berkeley: University of California Center for Research and Development in Higher Education, 1976.

16. Miller, James L., Jr. *State Budgeting for Higher Education: The Uses of Formulas and Cost Analysis*. Ann Arbor: Institute of Public Administration, The University of Michigan, 1964.

17. Mingle, James R. *Challenges of Retrenchment*. San Francisco: Jossey-Bass Publishers, 1981.

18. Sloan Commission on Government and Higher Education. *A Program for Renewed Partnership: The Report of the Sloan Commission on Government and Higher Education.* Cambridge, Massachusetts: Ballinger Publishing Company, 1980.

19. Stadtman, V.A. *Academic Adaptations: Higher Education Prepares for the 1980's and the 1990's*. San Francisco: Jossey-Bass Publishers, 1980.

Jane A. Lamb, Ed.D. is Associate Professor of one Department of Social and Health Resources at the University of Nevada Reno, Reno, Nevada 89557-0068.

Committee on Student Assessment

Proposed Final Report on Student Assessment

Frank Luth - Chair
Virginia Livingston
Andrew Kohen
Donald Corbin
Allen Lyndrup

April 10, 1986

Introduction

During the academic year 1984-1985, the administration and faculty of James Madison University (JMU) developed a plan entitled "Initiatives for Excellence and Accountability: A Five-Year Plan." The purpose of this plan was to put into motion a comprehensive review of the undergraduate academic program at JMU. In doing so, a master plan was envisioned as the outcome of that process. It is this plan that will set the academic direction at JMU for the remainder of this century.

A major goal of the five year plan is the development of a student outcomes assessment model. In order to accomplish this task, Dr. Frank Luth was assigned the responsibility of establishing a committee to investigate exemplary models and to develop a model for JMU. The following is the report of the Committee on Student Assessment.

During the academic year 1985-1986, James Madison University undertook a pilot study to determine alternate models of measuring student outcomes. The procedure was as follows:

In April 1985 the Administration of James Madison University, in co-operation with the State Council of Higher Education for Virginia (SCHEV), secured a grant from the State Legislature of Virginia to create a five year plan. One component of this activity was a plan to investigate student outcomes at JMU. During the month or May 1985, an extensive review of relevant literature led to the decision to investigate a variety of evaluation models.

During the early summer of 1985, four evaluation models were identified as worthy of investigation. Once the four models were identified it was decided that a pilot exercise would be conducted. Four volunteer programs were requested to develop a "game plan" as to how an assigned outcome model might be applied to their program. Four programs and respective investigation coordinators were identified.

A summary of this process and each of the four models is as follows:

Value-Added Model
Foreign Languages and Literatures
Dr. Donald Corbin

The Value-Added approach to assessment in higher education can measure, both cognitively and attitudinally, the amount of actual learning that is taking place at an institution. It can allow an institution to quantify the progress that is being made in any aspect of the programs in a university over a given period of time. Value-Added can be used to evaluate progress in general studies, in a major, in skills areas or the overall value being added to a student's knowledge, attitudes and career goals during the time that student is on campus. This method of assessment will permit comparisons of a given program to a national norm or to other programs or areas at the same university.

A. Advantages of Value-Added
   1) The model can demonstrate to students, faculty and administrators that progress is being made and that learning is taking place.
   2) It can indicate to the state legislature, SCHEV, parents and benefactors that the university is willing to hold itself accountable for the funds being used.
   3) It can draw students and faculty closer together as all cooperate to improve results. A student's chances of success in the outside world and the quality and prestige of the university can be enhanced at the same time.
   4) With adequate data and test scores at its disposal the institution will be in a position not only to assess the status quo but also to evaluate alternate courses of action for the future.

B. Disadvantages of Value-Added
   1) Some areas will be difficult to measure because they cannot be evaluated statistically; e.g., the acting ability of a theatre major.
   2) For some disciplines or areas there are no available test instruments to assess performance. In others existent instruments do not adequately measure the objectives set by the discipline or area.
   3) Faculty support is indispensable to implementation of Value-Added. If data, test results and surveys are used to accuse certain professors or to attempt to eliminate programs because of weaknesses, then the model will fail. Strategies must be found to identify and improve weaknesses rather than to eliminate weak areas.
   4) There is the danger of teacning solely to pass the GRE exam rather than teaching to meet set objectives.

Alverno College Model
College of Nursing
Ms. Virginia Livingston

Faculty at Alverno College, a private, liberal arts college in Milwaukee, Wisconsin, have created a curriculum based on abilities which will be integral to the students' successful adaptation as humans in a society experiencing accelerated change. Faculty at Alverno have identified abilities which are common to all traditional disciplines that are similar to identified common objectives at JMU. The abilities--effective communication, analytic capability, problem solving, valuing in a decision making context, effective social interaction, taking responsibility for the global environment, involvement in the contemporary world and aesthetic responsiveness--will enable students to change and grow, to assume varied roles across many settings througnout the lifetime of their careers.

The learning of the eight abilities is organized in a progressive pattern. The faculty have broken down each ability into six developmental levels. Development of the first four levels of all of the abilities is validated by each student in the general education sequence. Levels five and six of only those abilities pertinent to the major are assessed in the junior and senior years.

Validation is accomplished through a university-wide, individualized assessment process: the hallmark of the Alverno approach. It is a system of measuring each individual student's progress in developing the abilities and relies on assessment instruments created and used in the classroom. Instructors have developed assessment instruments which will validate the student's ability to integrate the knowledge learned in each class with a certain level of one of the abilities. At each level instruments are developed which use course content as context.

Additional features of the Alverno Model include:
1. Baseline assessment of entering freshmen.
2. A progressive videotape record of each student's communication ability.
3. An Assessment Center staffed with trained assessors.
4. An Office of Evaluation and Research.
5. An Instructional Services Center to assist with student learning.

Strengths of the Alverno Model wnich could be applied at JMU include:
1. Creation ot an Assessment Center to assist faculty with student assessment and to provide faculty with education in the assessment process.
2. Development of an Assessment Center staff trained to carry out faculty designed assessments.
3. Freshman assessment in the area of communication and group interaction in addition to traditional testing of cognitive ability. Videotaping freshman communication performance.

4. Creation of additional assessment instruments which will require students to validate speaking ability in selected general education courses and courses in the major.
5. Writing across the curriculum.
6. Creation of discrete and focused instruments to assess the JMU common objectives.
7. A growing emphasis on creating learning situations which require production rather than recognition measures by the students.
8. Increased emphasis on students' self assessment--assurance of future autonomy and quality.

Potential Concerns of the Alverno Model:

1. Integration of transfer students.
2. Expense. Creation of support services to implement this model.

Tennessee Performance Model
Economics
Dr. Andrew Kohen

The Tennessee Performance Funding Model is essentially a set of five quality standards created to stimulate the public institutions in Tennessee to become involved in evaluating educational outcomes. The extent to which the standards are met by an institution determines its eligibility for supplemental funding from the state government. The five elements of the model are as follows:

1. Accreditation of programs by external agencies.
2. (External) peer review of programs and/or administration of a comprehensive field examination to student majors.
3. Measurement of the "value added" by the general education component of the curriculum (as gauged by the ACT COMP examination).
4. Assessment of program quality from opinion surveys of students, alumni, employers and community members.
5. Implementation of a plan for instructional improvement based on the diagnostic findings from the first four elements.

The Tennessee Performance Funding Model Applied to JMU.

Obviously, the elements of the model could be applied to James Madison University because it was created to embrace the operation of an entire institution of higher education. However, because its scope is that expansive, there are dimensions of it which are definitionally inapplicable to the specific program of a Bachelor of Science Degree major in Economics. The first element is irrelevant to this program because there is no pertinent external accrediting agency and the third is irrelevant because its dimensions surpass the operation of any single department.

The model does encompass entrance testing of rising freshmen, in the third element, in order to establish the "base" to which the general studies curriculum is presumed to "add value." Administration of the COMP exam would also be a method of assessing student acquisition of some of those attributes identitied at JMU as being common objectives. In addition, the model stipulates alternative modes of assessing student outcomes within the major field of study, as found in the second and fifth elements of the model. In particular, there are three modes. One is external peer review. A second is administration of an internally or externally created comprehensive examination to student majors, and comparison of the results with either performance of students in other universities, or the historical performance of students in the same university, or both. A third is collection from alumni, current student majors and employers of attitude data which are pointed at assessing the "quality" of the major field program.

Strength of the Tennessee Performance Funding Model

A significant strength of this model is its flexibiiity in accomodating a variety of measurement devices and methods to assess instructional quality. Another strength of the model appears to be its comprehensiveness insofar as it stipulates that the quality of an institution be assessed with reference to its general education curriculum along with the major field curricula. It also accomodates both internal self-study and external expert review. It also is holistic in its approach in that it mandates collection of qualitative and quantitative data.

Still an additional strength of the model is that its campus-wide adoption would require each departmental faculty to confront its institutional programs directly, explicitly identify the expected student outcomes and decide whether to hold itself accountable for improvements over time or to some externally determined standard or both. Finally, a strong aspect of the model is that it suggests that rewards should be provided for those programs which implement changes derived from the diagnoses emanating from the various measurements. Absent the stipulation that such rewards are to be part of the system, the assessment activity becomes an empty one which, at best, establisnes "bragging rights."

Weakness of the Tennessee Performance Funding Model

Some aspects of the model may also be viewed as weaknesses. The flexibility and nonspecificity of the model with respect to measurement devices and methods leaves open the possibility of noncomparability of outcome results between departments within an institution and/or across institutions. Failing the first comparison, the assessment efforts would yield no guidance on possible intra-institutional (re)allocation of resources, irrespective of whether the goal of such reallocation were rewarding excellence or remediation in pursuit of excellence. Failing the second comparison, the assessment effort would subject itself to no external validation criteria or norms and might be subject to the criticism of being self-serving and parochial.

Another weakness of the model is its failure to link explicitly the general education and major field assessments. The lack of specificity of the model also leaves open the possibility that each (many) department (s) within an institution will waste time in "reinvention of the wheel." Another shortcoming of the model is its failure to address the two alternative interpretations of the concept of "value added" in higher education. It focuses on the more conventional interpretation, namely a quantitative measurement of cognitive gains by students compared to gains by earlier cohorts of students in the same program (institution) or compared to gains over the same time frame by students in the same program at other institutions (individually or on average). Nowhere is there reference to the alternative of measuring "value added" in terms of student gains compared to potential student gains.

Discrepancy Evaluation Model
Theatre
Mr. Allen Lyndrup

The Discrepancy Evaluation Model (DEM) is a pure evaluation model, it can be structured to define, then assess inputs, processess and outputs of any organization or program. Evaluation of a program illustrates discrepancies in it, modification of the program or revision of the program standards should minimize those discrepancies over time. Discrepancy evaluation differs from other models in that it starts with a program design, incorporating evaluation only after the program has been carefully described.

The DEM borrows a systems analysis technique to illustrate the organization of a program. Components in a unit are laid out in Networks that show interrelationships and interdependencies. Each component in a Network is defined in terms of its inputs or resources, its processes or how it sets about to do its work, and its outputs or the goals and objectives of the unit. Networks, with their Input-Process-Output documents (IPO's) are first establisned at a very basic level, each component is then broken down into its own Network with its IPO until the appropriate level of complexity is reached. It is possible to carry out this process to different levels for different parts of each program, depending on perceived evaluation needs.

As the Program Design is developed, certain evaluation needs will be identified. Other outside evaluation needs--accreditation standards, for example--also are defined early in the process. These perceived needs and the strategy to provide the information to satisfy them form the basis for the Evaluation Plan. In the Evaluation Plan, the Evaluation Design is first prepared to provide an overview of the conceptual structure of the intended evaluation effort. Evaluation Questions are then proposed and analyzed to reduce broad concepts to specific, measurable variables. Finally, an Evaluation Workplan is proposed, laying out assessment instruments to be used, how and when they will be administered and how the resulting data will be analyzed and reported.

After the Evaluation Design is carried out, the program staff uses it to determine discrepancies that exist. Negative discrepancies can be resolved by modifying the program to compensate for them or by adjusting unrealistically high performance standards. Positive discrepancies can be used as evidence of excellence in a program component, or might also indicate unnecessarily low performance standards. As more data are gathered, each program and its standards are adjusted to remove or minimize discrepancies.

There are several strengths and, conversely, several weaknesses perceived in the Discrepancy Evaluation Model.

Model Strengths:

Flexibility. The Discrepancy Evaluation Model is flexible enough to evaluate the inputs, processes and outputs of traditional, nontraditional, interdisciplinary and out-of-classroom programs.

Strong relationship between design and evaluation. The staff of a program must clearly define how a program functions before it begins to determine what it is going to evaluate. *This process alone may well be worth the effort of undertaking a Discrepancy Evaluation.*

Effect on curriculum change. Shortcomings within a program will point the way for curriculum modifications.

Explicit, open and understandable. Jargon is kept to a minimum, definitions of terms are relatively clear and precise, and procedures are well defined.

Similarity in language and framework. Anyone having prepared a Discrepancy Evaluation will be able to understand the documents of other programs, the pictorial nature of the documents should make them understandable to individuals unfamiliar with the intricacies of the model.

Use of existing forms of evaluation. Current evaluation processes can be worked into the evaluation design.

Emphasizes expertise of program staff. The staff becomes expert in determining standards and assessment measures, interpreting results and reshaping the program to alleviate discrepancies.

Model Weaknesses:

Evaluation determined on a program-by-program basis. It is desirable to involve individual staffs of programs in setting standards and assessing performance, and it is also desirable for assessment to be tailored to individual programs. However, since standards are set and measured within programs, it will be difficult to use the data in any sort of comparison between programs. If the goal of assessment is to help individual programs improve, then this is no weakness. If the goal of assessment is to make comparisons between programs, then outside controls will have to be imposed on the assessment process.

Possible need for outside validation. As noted above, program design and assessment are determined by the staff of each individual program. Outside experts in Discrepancy Evaluation and in the subject area might need to review the Program and Evaluation Design in order to validate it.

Can be perceived as complicated and time consuming.

Academic prejudice against "Education School" models.

Seems better suited for smaller units of an individual program.

The biggest weakness in the Discrepancy Evaluation Model is the perception on the part of the faculty that the DEM is going to bog them down in unnecessary paperwork and keep them from their primary function, teaching. Convincing the faculty that assessment is a part of their responsibilities and can provide them valuable information to make their work more rewarding will be a major undertaking for the administration of the university.

Development of JMU Model

With the pilot activities concluded, the committee turned its attention to developing a model for JMU. In doing so, the committee looked to the need for creating a balanced approach to assessment. It was evident that a number of concerns would have to be addressed in the development of a model. A major objective of the committee was to keep those factors in balance. Areas of concern were as follows:

1. Curriculum Advancement. Regardless of the model adopted, a major objective of the process must be the improvement of the curriculum. The model, therefore, must provide information which would be of value in making curriculum decisions.

2. Internal Curriculum Validity. The model developed must allow faculty to determine what they teach and how they teach. At the same time it must validate that they teach what they say they teach.

3. External Curriculum Validity. For certain programs or areas of interest it may be of value to compare JMU to institutions with similar missions. In such instances, instruments of a standardized nature will be necessary.

4. Political Considerations. It is apparent from the literature that many state legislatures have become directly and forcefully involved in setting assessment priorities in higher education. A review of various

approaches leads one to believe that this is only going to become more pervasive, with long term implications for higher education. Obviously, the model adopted by JMU would have to address any concerns generated by our own state legislature.

5. Funding Considerations. Assessment can be extremely expensive. The committee was cognizant of the need to develop a comprehensive but at the same time a cost-effective model.

6. Academic Freedom. Of vital concern to the faculty was the need to ensure that academic freedom was protected throughout the process. The committee was particularly concerned with identifying procedures which would ensure faculty the freedom to determine not only course objectives but also assessment procedures to be employed.

7. Resources
   A. Faculty time - Assessment activities require time and energy. Any model developed should not put an unusual and unreasonable burden on the faculty as a whole. Current efforts are underway to secure resources to fund part-time faculty to support all aspects of the Five-Year Plan.
   B. Student time - A successful assessment program will require meaningful student participation. Careful attention in developing a model will have to be directed toward balancing the need for data and the willingness of students to participate.
   C. Staff time - Clearly, the development of a comprehensive student assessment model calls for the management of a significant amount of data. A careful balance will have to be kept between the need to store and organize data and staff resources.

Tentative JMU Model

As a result of the pilot programs as well as the input from many of the other committees participating in the Five-Year Plan, a seven dimension model has been developed. The components are:

1. General Studies. This dimension will provide an assessment of a general nature regarding the overall impact of the general studies program. The ACT-COMP assessment procedure will be piloted in the spring of 1986 to determine its appropriateness for JMU. The pilot exercise will be conducted by the Liberal Studies Committee using a random sample of seniors.

2. Common Objectives. A second dimension of interest is the area of common objectives. At the present time, the Common

Objectives Committee has tentatively identified a list of objectives and is turning its attention to assessment methods.

3. Functional Skills. The third dimension is the area of functional skills. During 1986-1987, functional skills areas will be defined and pilot exercises will be conducted to determine appropriate assessment procedures and minimum achievement expectations. Initially, a sampling procedure will he used for JMU as a whole. If results are acceptable, subsequent assessments will focus on students who might be considered "at risk" because of preparation deficiences.

4. Quality of Effort. To what degree are students challenged at JMU? The Degree to Challenge Committee has been investigating this question actively for the past year. To determine the answers to this question, a pilot exercise has been conducted this spring. Faculty perceptions will also help contribute to the data base. An intra-university sampling procedure was used in this procedure.

5. Major. Perhaps the area of academic assessment which will have the most direct impact on programs is the assessment to be conducted in the major. It is anticipated that this area of assessment will take five years to implement using a phase-in procedure. Each department will have the responsibility to determine objectives in each major and instruments to be used. It is projected that extensive technical assistance will be available in this area of the project.
6. Non-Cognitve. During the next year a cross-section of the JMU community will take an extensive look at non-cognitive areas of interest. This particular area of investigation is included in order to bring into perspective the fact that an undergraduate program must address more than just cognitive experiences.

7. Alumni. The last dimension to be included is follow up of alumni. It is anticipated that alumni will be surveyed with respect to general university concerns as well as issues specific to a given major.

The seven dimensions will be implemented in an intra-university/intra-departmental format. The reader is directed to Table I for a breakdown of activities and responsibilities. It is anticipated that the intra-university assessment activities will either be conducted or coordinated by a proposed Office of Student Assessment. Intra-departmental assessments in the major will be the responsibility of each department. Follow up of alumni will be a collaborative effort between the proposed Office of Student Assessment and departments. A tentative schedule of assessment procedures for the undergraduate sequence is provided in Table II.

During the spring of 1986 the General Assembly of Virginia accepted the recommendations of Senate Document No. 14 entitled "The Measurement of Student Achievement and the Assurance of Quality in Virginia Higher Education." The specific recommendations proposed by the Council ot Higher Education for Virginia are included in Appendix I. A review of these new requirements will reveal that the proposed model for JMU meets or exceeds all six requirements.

In conclusion, the Committee on Student Assessment wishes to take this opportunity to strongly recommend support for this student assessment model. While comprehensive, it is efficient. At the same time the control and direction of specific assessments such as the ones to be conducted in the major have been left to the discretion of the faculty. There is a clear promise of an ongoing support system as a result of new funding from several sources. Rather than being a system of assessment imposed upon the university from an external agency, this model represents the best thinking of a cross-section of the JMU community.

The Committee for Assessment would like to recognize the many individuals who have contributed to this process. This is particularly so in the instance of the outstanding consultants who have visited the campus and members of the other committees of the Five-Year Plan. A special expression of appreciation is extended to Ms. Margaret Wichael and Mrs. Barbara Shifflett, secretaries and Ms. Lou Combs, graduate assistant for their dedication to the objectives of this project. Finally the committee would like to acknowledge the ongoing support of Dr. Russell G. Warren, Vice President for Academic Affairs, who made this investigation possible.

TABLE I
JMU MODEL

DIMENSIONS OF ASSESSMENT

Intra-University - Office of Student Assessment

(1) General Studies - ACT COMP - Other

(2) Objectives Across the Curriculum - ACT COMP - Other

(3) Functional Skills

(4) Quality of Effort (Degree of Student Challenge) - Pace Instrument

(5) Non-cognitive - Affective - Psychomotor - Other

(6) Alumni - Survey - Cognitive/Affective/Other

Intra-Departmental - College/Departmental

(1) Major - Standardized - Criterion

(2) Alumni - Survey

(3) Other

TABLE II
TENTATIVE MODEL

JMU Student Outcomes Assessment Model

Rising Freshmen	Freshmen	Sophmore	Junior	Senior	Alumni
Functional Skill Assessment (Selected Students)	Functional Skill Assessment (Selected Students)	Functional Skill Assessment (Selected Students)			Follow-up
General Studies Assessment (COMP)	General Studies Assessment (COMP)	General Studies Assessment (COMP)		General Studies Assessment (COMP)	1 Year
					5 years
Common Course Objectives Assessment	Common Course Objectives Assessment	Common Course Objectives Assessment		Common Course Objectives Assessment	- Cognitive
Other Non-Cognitive Assessments		Other Non-cognitive Assessments		Other Non-cognitive Assessments	- Affective
	Quality of Effort Assessment (Pace Instrument)	Quality of Effort Assessment (Pace Insturment)		Quality of Effort Assessment (Pace Instrument)	- Other
		Entry Level Assessment in the Major (Selected Majors)		Comprehensive Assessment in the Major	
				Exiting Senior Interview	

--10 Year Assessment Period--------------------------------------

Draft- January 1986

APPENDIX I

Report of Study Conducted By The Council of Higher Education For Virginia

The Measurement of Student Achievement
and the Assurance of Quality
in Virginia Higher Education

To The Governor and The General Assembly of Virginia

Senate Document No. 14

Recommendations

Recommendation 1. That the academic relationship between secondary and higher education be strengthened:

(a) By developing programs such as the Ohio Board of Regents' Early Testing Program to help high school students prepare for college and the Minnesota early admissions program to reward those who demonstrate an ability to do college-level work; and

(b) By providing reports from colleges and universities that tell the high schools how well their former students are doing in college.

Recommendation 2: That all state-supported institutions of higher education establish procedures and programs to measure student achievement. These programs should:

(a) Derive from institutional initiatives, recognizing the diversity of Virginia's public collges and universities, the tradition of institutional autonomy, and the capacity of faculty and administrators to identify their own problems and solve them creatively,

(b) Be consistent with each institution's mission and educational objectives;

(c) Bear a direct relationship to teaching and learning in the classroom, enabling faculty to use the results to address student deficiencies, evaluate and improve the curriculum, and develop better teaching techniques;

(d) Involve faculty in setting the standards of achievement, selecting the measurement instruments and analyzing the results;

(e) Consider the relative importance of both assessment to determine student attainment as measured by an absolute standard and assessment of student growth in learning attributable to the influence of the institution;

(f) Follow student progress through the curriculum, as appropriate, with consideration of achievement measures (1) at transition points to ensure student readiness to proceed, (2) upon completion of the major, and (3) at graduation or on leaving the institution; and

(g) Include follow-up of graduates through employer surveys, studies of participation rates in further education and alumni reports of career progress.

Recommendation 3: That institutions administer tests to determine the entry-level skills of students whose past performance, as defined by high school grades or Scholastic Aptitude Test scores, indicates they might have difficulty doing college-level work; and that each institution identify a minimum threshold of achievement to qualify for college degree-credit courses.

Recommendation 4: That institutions with students whose skills fall below the threshold established for college-level work provide remedial education to maintain access while improving the quality of students' performance prior to full participation in degree credit courses. Where possible, remediation for students at four-year institutions should be arranged through agreements with community colleges. No credit toward a degree should be awarded for remedial work.

Recommendation 5: That an advisory committee to the Council of Higher Education be established to develop guidelines for designing good assessment programs, to assist the institutions on request to develop the programs, and to advise the Council on progress in this area.

Recommendation 6: That the state-supported colleges and universities submit annual reports of progress in developing their assessment programs and concrete, non-anecdotal and quantifiable information on student achievement to the Council of Higher Education. The reports should include information about the achievement of transfer students from the community colleges enrolled in four-year colleges and universities and about the performance of professional program graduates on licensing and certification examinations. The Council should publish results of the assessment programs and reports of other actions to strengthen educational quality in its biennial revisions to the Virginia Plan for Higher Education.

Frank Luth chaired the Committee on Student Assessment at James Madison University, the Committee members included Virginia Livingston, Andrew Kohen, Donald Corbin and Allen Lyndrup.

CREATING USABLE PROGRAM EVALUATION

James H. McBath

Several years ago the graduate dean of a research university in the South stated bluntly, "Take it from me, under any circumstances review is traumatic." The thrust of his remarks at a Council of Graduate Schools annual meeting (Bollier, 1974) was that needed change could be thwarted by negative attitudes toward academic program review and evaluation: "Faculties who feel themselves under attack will hardly cooperate with good grace in making even those adjustments and changes which hard facts dictate. We can legislate such adjustments and changes, of course, but legislation by fiat leaves much to be desired if we desire a viable institution."

Assessment is rarely welcomed by those undergoing review, except possibly by departments or programs that are supremely self-assured. Even in prosperous times, review can be traumatic. During periods of retrenchment, fears abound that evaluation may lead to the termination of dismemberment of programs, the displacement of faculty and staff, and the erosion of academic quality and reputation. For the period of review, any program feels itself at risk. Of course, the sources of internal resistance to program evaluation are understandable. People and institutions have a stake in preserving a status quo they helped to create. Even a favorable review is likely to cause some change; negative assessments can result in the recommendation of significant modifications or loss of resources. In extreme cases, program termination is a possible outcome. The whole point of an assessment is its usefulness in educational decision making.

Reception of Evaluation

In a sense, every academic evaluation is an argument for change; rarely, if ever, does a report conclude that the program is optimum as it presently stands. The recommendations rest upon evidence that is testable and criteria that provide an etiology of the inferences. Conclusions of evaluators may be compared with those of others and their reasoning is open to the scrutiny of informed and interested readers. Indeed, it is helpful if the several audiences can see clearly the comparative advantages of the recommended actions over the principal alternatives. Certainly the strengths and assets of a current program should be weighed against those of a prospective modified program. Then if the evaluation goes one step further and reveals the educational rationale for judgment, it acquires even more significance as an argument for change because readers are invited to confront the assessment with greater precision.

The most compelling arguments, besides satisfying rigorous standards of proof, will engage the attitudes of those affected and motivate them to accept what the logic of the matter seems to dictate. This is no simple task. But consultants should remember that whether a review is voluntary or mandated, a vital audience of the argument for change is the faculty and administrators of the academic unit whose program is undergoing review.

Inevitably, resource decisions will determine the shape of the institution. Such decisions cannot be made wisely without information, understanding, and mutual trust. Program review, properly conducted, can contribute to each of these requisites. Arns and Poland (1980) have suggested that the most important outcome of assessment "may well be the strengthening of confidence in decisions, their evolution and their implementation. None of these benefits are inherent in the term 'program review.' They all depend on the way it is done" (p. 283). Yet in the dialogue on reasons and methods for conducting evaluation, relatively little attention is paid to the reception of evaluation by program participants. The oversight is unfortunate. The ultimate success of program review depends on the readiness of those most directly involved to implement measures that will improve academic quality. Usable program evaluation consciously seeks acceptance, if not agreement, by the program under review. Its concern is what happens at the institution after the review is completed.

Communication of Purposes and Standards

The likelihood of creating usable program evaluation is enhanced if program participants understand both the purposes and standards of review. State agencies, administrators, consultants, and faculty should have a common perception of the objectives and potential outcomes of assessment. To do otherwise is to risk a general lessening of confidence in the evaluation process.

The experience of one large academic unit of a comprehensive state university is instructive (McGlone, 1984). The Board of Trustees of State Institutions of Higher Learning announced that all degree programs in all eight public universities would be reviewed. This meant more than 750 reviews, including forty-seven in the College of Arts and Sciences. Upon completion of the review process, the presidents of the eight universities were summoned to a press conference at the Board office. There the Board president announced that in executive sessions during the previous weekend, the Board had developed and adopted a mission statement assigning responsibility for certain programs to certain institutions and to eliminate some programs and offerings based on that statement. Decisions often were counter to recommendations of the program review; the recommendations themselves were misinterpreted or misstated. Inevitably, the integrity of the process was called into question.

Several weeks later a special committee of the National Association of State Universities and Land-Grant Institutions sent a carefully worded statement to the Trustees that concluded, "To be successful, such review processes should be conducted with care and integrity so as to encourage the thoughtful advice of the faculty and administration responsible for a given program. Consultants should be made fully aware of the intent of the review so that their work is not compromised in the process." The dean of the college then offered this advice (McGlone, 1984):

> Before you are reviewed, be clear and candid about the purpose of the review. If the review is a mandated one, try to get the dean or vice president or president or Board to be similarly clear and candid. The best way to do this, the only way perhaps, is to get the appropriate parties to make both written and oral statements about what the review is supposed to accomplish and what is not intended by the exercise. The problem we had at my institution is that there were purposes intended which were never shared with the consultants or the persons in the programs under review. (p. 23)

Usable program evaluation also is more likely to occur if program participants understand the standards for assessment. Criteria must be reasonable, fair, and comprehensive if they are to provide durable bases for judgment. For example, David Longanecker (1979) of the Congressional Budget Office has proposed three conditions as important for developing dependable measures for quality assessment:

1. Using data that are reliable and readily available. Information in the public domain may be acquired unobtrusively and at modest expense. Moreover, data that are commonly collected from all programs in a field allow easier and more reliable comparison of programs.

2. Creating measures that distinguish between levels of quality along the entire continuum of programs from those of very high quality to those of unacceptable quality. While it may be interesting, and possibly important, to know who is on top, it is equally important to know the relative quality of other programs. Indeed, a large share of assessment activity focuses on programs of lesser or moderate quality.

3. Devising techniques used to assess quality that distinguish between inputs into the quality function and outcomes or products of the educational process. The outcomes provide an index to how well the process is working. While this is a critical factor in measuring quality, it is not, by itself, a sufficient measure of quality. We need also to examine the various components of the educational process and ascertain how they contribute to the outcomes.

The Usable Report

The written advisory products of evaluation, besides a clear communication of purpose and quality measures, should be framed with the clientele audience in mind. Here it is well to remember that while many assessment activities are intended for use by departments and individual institutions, statewide reviews attempt to provide information for lay board members, state legislators, and administrators who are called upon to make decisions about the commitment or continued allocation of resources. But whatever the purpose or sponsorship of evaluation, the prime users of the recommendations are those responsible for administering and executing the program on campus. For this audience, the usable evaluation report possesses several commonsense characteristics.

First, the report is audience-centered. From the outset, an audience-centered report is a collaborative enterprise. It is not enough to involve audiences in evaluation simply so they will ''feel they had a part in the process"--however true that may be--but also to improve the evaluation. Feedback relationships are essential to effective evaluation. Indeed, appraisal itself is formal, systematic feedback. Feedback not only increases the accuracy of information, it also increases the confidence of both evaluation team and audience. On the other hand, zero feedback fosters doubts in both parties. One way communication runs the risk of producing defensiveness that can impair the openness required for acceptable evaluation. However, a dialogue can reveal facts, interpretations, or judgments previously not perceived. New ideas may emerge from the pooling of information and the exchange of views. Errors may be exposed, since lines of thought are subjected to tests of their proof and consistency. Evaluators themselves may discover more productive patterns of analysis. The most effective evaluation results from interaction of evaluators and program participants.

An audience-centered report is sensitive to the requisities for academic change. Because every modification in academe has a sort of chain reaction with other academic units, revised programs should be compatible with existing structures and other programs. The pragmatic report also permits sequential implementation of recommendations. Even reasonable recommendations may be resisted if the report appears to insist that program improvements must be made all at once. Perhaps some steps can be taken immediately, and others later. Evaluation reports should not overwhelm their audiences. If they do, the probable result will be negative reactions to the whole report. When resistance is foreseen, advancing one step at a time holds the greatest likelihood for creating a climate of acceptance. Incremental improvement is preferable to none at all.

Second, the evaluation must be grounded in reliable academic information. Evaluators must know the nature and dimensions of the academic field under review. Those possessing comprehensive knowledge are rare; most academicians live within their own specialties. Evaluators, however, must be keen students of their fields and aware of its varieties and possibilities. Besides expertise in curriculum, they should have up-to-date information about the academic and nonacademic Placement of degree holders. The evaluator in my field, for example, can examine national employment patterns for graduates over the past dozen years. The published record displays career experience in terms of specialty, geographic location, and occupational trends. It would be foolhardy to recommend curricular change without reference to the opportunities for candidates upon graduation. Studies within individual disciplines can also identify the prospects for nonacademic employment and the kinds of preparation indicated.

Program evaluators should know the educational cosmos they are evaluating. For example, we are familiar with the classification systems in use since 1971: Standard Terminology for Curriculum and Instruction in Local and State School Systems, often referred to as the Handbook VI Taxonomy, and A Taxonomy of Instructional Programs in Higher Education, commonly known as the HEGIS Taxonomy. In 1975 the National Center for Education Statistics commissioned the development of a new classification system. Twenty-one subject-matter committees advised on the content of the document, including conceptualizations of fields of knowledge and definitions for the subject-matter entries. This later taxonomy, A Classification of Educational Subject Matter, published in 1978 by the National Center for Education Statistics, is a valuable reference for evaluators.

Reinforcing the essential academic dimensions of assessment are guidelines found in the work of the Joint Committee on Standards for Educational Evaluation (1981), whose five"year study was released in 1980. Written, reviewed, and field-tested by more than 200 educators and evaluators, the Standards are a compendium of principles and practices intended to result in evaluations that are useful, feasible, fair, ethical, and technically sound.

Finally, evaluations should take advantage of the relatively small but growing body of information on quality assessment application. While much of the research has dealt with doctoral programs, findings have been extended to master's programs. It is likely that some findings have a special relevance to master's programs. Because master's candidates are new to graduate work, their reactions to the process dimension--the nature and quality of the experiences of those who are part of the program--may be particularly germane. Process indicators include such items as quality of teaching, climate for learning, faculty members' concern for students, degree requirements, and the like (Clark, 1976). These judgments, perceptual in nature, are derived from program Participants rather than being based on discernible characteristics. Students in general are less concerned about the level of faculty research activity than they are about the environment for learning.

The third characteristic of acceptable program assessment is that it capitalizes on the communication potential of the evaluative process. Evaluators have a unique access to an institution's communication network. They can interact with faculty, students, deans, administrative staff, and, sometimes, alumni. Relevant information, current and historical, is available to them. Their audiences are attentive and interested; all have a stake in the outcomes of program review. Evaluators can become key links in exchanges that will foster understanding and have residual benefit to the university.

A major function of evaluation should be the facilitation of intercommunication. Even where faculties and administrators have ideas about improving programs, they may have difficulty in talking with each other about mutual problems and working through issues cooperatively. Clearly, one of the major functions that evaluators can serve is to facilitate communication within the department and between department and administration.

Late in the assessment period, evaluators often meet again with administrators to discuss initial impressions, to answer questions, and to agree on a form of the final report. These culminating sessions can be the most important of the entire review process. In face-to-face conversation with administrators of the institution, evaluators have the chance to convey their impressions about the field and about the programs being reviewed. Evaluators can speak frankly about sensitive issues that may not appear in the written report: tension between junior and senior faculty, student grievances, forseeable difficulties in promotion and tenure, personnel conflicts, and the like. Administrators want ideas on how to interpret and approach such problems. They are equally interested in learning about program accomplishments that previously were overlooked. Many of the topics may also be discussed with members of the internal review committee as well as with departmental officers. Thus the evaluation process itself can serve as constructive communication within the institution.

So far we have considered several tactics for advancing the acceptance of program evaluation. Of course there are other ways, some outside the university. The disciplinary associations themselves, besides accumulating and analyzing departmental data, might well create a broader understanding of program evaluation. In my own discipline, for example, the Association for Communication Administration periodically conducts conferences and workshops on evaluation. Representatives from 80 to 100 institutions present papers, engage in discussions, and hear resource speakers discuss current research in program evaluation. Conferees have reported that the meetings are useful in achieving these objectives: (1) to see what others are doing in evaluation, (2) to provide familiarization with a range of issues and approaches to evaluation, (3) to give exposure to arguments for and against various evaluation procedures, and (4) to suggest specific evaluation procedures applicable to one's own institution.

One can never know in advance if a program evaluation will result in modest improvement, a complete overhaul of curriculum, or a ringing endorsement of the status quo. But a completed assessment must be accepted if it is to exert any influence at all. It must be a cogent statement of findings and recommendations, an argument in which decision makers have confidence. Clearly, how the review is conducted is at least as important as why it is done, and the creation of usable assessment depends on the quality of the process and its presentation.

References

Arns, Robert C. and William Poland. "Changing the University through Program Review." Journal of Higher Education, 1980, 51, 268-284.

Bollier, E. Phillip. "Internal Evaluation: A Challenge for Change." Council of Graduate Schools of the United States, Proceedings of the Fourteenth Annual Meeting, December 1974, 236.

Clark, Mary Jo. "Multiple Measures Assess Doctoral Programs." Findings (Educational Testing Service), 1976, 3, 3.

Joint Committee on Standards for Educational Evaluation. Standards for Evaluation of Educational Programs, Projects, and Materials. New York: McGraw-Hill, 1981.

Longanecker, David A. "Assessing Quality in Graduate Education: A Multivariate Approach." Proceedings of the Conference on the Assessment of Quality in Master's Programs. University of Maryland, 1979, 91-100.

McGlone, Edward L. "Program Evaluation and Elimination: A Case Study." ACA Bulletin, 1984, 50, 19-23.

James H. McBath is Professor of Communication Arts and Sciences at the University of Southern California, Los Angeles, California 90089-1694

STATE CONTROL OF HIGHER EDUCATION IN MICHIGAN: A NEW SCENARIO

by

T. Harry McKinney

The final report of the Governor's Commission on the Future of Higher Education in Michigan represents an important step in statewide efforts to review and improve higher education. The Commission highlights many of the issues and offers numerous recommendations calling for increased financial support.

At the same time, however, the Commission includes several recommendations that would lead to greater centralization of control over public institutions of higher education, particularly state colleges and universities. This would be done through the state budgetary process, with the Department of Management and Budget and the Department of Education playing more important roles.

What are the recommendations? How would they lead to more state control? And what are some of the conclusions that might be drawn from them? These are a few of the questions that need to be considered as efforts to implement the recommendations are made.

Historical background

To understand these recommendations, it is important to keep in mind that Michigan has one of the most decentralized systems of higher education in the nation. Each of the public institutions has its own governing board with constitutional responsibility for general supervision of the institution as well as control and direction of all expenditures from institutional funds.

Constitutional provisions also give the State Board of Education responsibility for planning and coordination of all public education, including higher education, and for advising the Legislature on financial requirement. However, a decision of the Michigan Supreme Court in 1975 made it clear that the Board's role in planning and coordination of higher education is advisory only. For this and other reasons, the State Board of Education has had little or no influence on major policy decisions related to higher education for many years.

Instead, representatives of the various public institutions have been relatively free to approach the Governor and the Legislature for funds needed to introduce and support programs viewed as being essential to or consistent with institutional goals. This has led to a situation which institutional representatives generally view as being directly related to institutional vitality, programmatic flexibility, and education quality.

Recommendations by the Commission

The Commission has taken the position, however, that educational quality can only be maintained if more decisions about educational matters are made at the state level. Recommendations that would seem to be most important in this regard are those related to (1) identification of institutional roles and missions, (2) development of a comprehensive higher education database, (3) adoption of policy-based funding formulas, (4) provision of better mechanisms for program review and approval, and (5) promotion of program efficiency.

For each of these topics, and others throughout the report, the Commission makes a general recommendation followed by one or more specific recommendations. The general recommendations indicate what should be done. Specific recommendations reveal the agency or agencies responsible for related actions.

Institutional roles and missions. The general recommendation concerning institutional roles and missions states that these should be "carefully and explicitly defined" and that "future state decisions to fund or not fund educational programs be based on these role and mission statements."

To support this recommendation, the Commission refers to financial constraints, declining enrollments, and shifting demands for educational programs. It also contends that distinctions among institutions would not imply that some are more important than others, that its own distinctions are supported by the facts, and that individual institutions would find it difficult to make the distinctions themselves.

The Commission then recommends specifically that "a classification system be adopted by state policymakers as the framework for state funding decisions." This is followed by a description of the Commission's classification which it later recommends for adoption. In its classification, the Commission divides public baccalaureate institutions into five categories: public community colleges, technical colleges, regional state colleges, general state universities, and nationally-recognized research universities. Within each category, the Commission specifies the general nature of programs to be offered, degrees to be awarded, and areas to be served. It also specifies the institutions to be included.

Higher education database. The general recommendation calling for development of a comprehensive higher education database specifies that it should include all institutions of higher education, public and private, and should also include data realted to employment opportunities on the one hand and placement of graduates on the other.

To support this recommendation, the Commission contends that "higher education decisionmaking in Michigan must be done on a systemwide basis if it is to be effective" and the "higher education policymakers lack a continuous comprehensive source of data for timely discussion of complex issues." The Commission then calls for "expeditious" development of a comprehensive database to be administered by the Michigan Department of Education and developed in cooperation with the Departments of Commerce, Laobr, Management and Budget, and the legislative fiscal agencies.

Policy-based funding formulas. The general recommendation related to policy-based funding formulas specifies that these are needed "to achieve state goals and recognize enrollment changes and institutional cost variations."

To support this recommendation, the Commission refers to limitations of current approaches to state funding of higher education in Michigan and notes with approval recent efforts to develop a policy-based formula for community colleges. It then recommends immediate action by the Executive Office and the Department of Management and Budget to create a task force including representatives from the Presidents Council of State Colleges and Universities, the Legislature, and the Department of Education for the purpose of revising a funding model for state colleges and universities used in the alte 1970's.

Program review and approval. The general recommendation related to mechanisms for program review and approval calls for the state to adopt the proposed institutional role and mission classification system and develop "a more systematic budget process" for higher education that would include "a program review and approval process."

To support this recommendation, the Commission takes the position that "allocation of educational resources by clear policy goals is the most appropriate approach to budgeting" and the "leadership of the Governor and the Legislature is pivotal to this effort." The Commission states further that in its view the budget process is "the most workable means for focusing priorities."

The Commission then makes three specific recommendations related to the state budgetary process. The first would give the Department of Management and Budget and the Department of Education responsibility for developing the new budgetary process for higher education at the direction of the Governor. The second calls for the Governor and the legislature to establish procedures for review and approval of programs within that process with the understanding that programs found to be inconsistent with specified criteria would not be funded. The third calls for appointment of a new commission after three budget cycles to assess the result.

Program efficiency. The general recommendation related to program efficiency calls for implementation of the various measures described above combined with development of criteria to be used in situations where institutional mergers or closures might be needed.

To support this recommendation, the Commission contends that a large number of undergraduate and graduate programs still in operation are not needed. It then recommends a plan for a "stronger, leaner, and more efficient system" that would result in substantial savings related to capital expenditures and operating costs. The plan would focus on several elements including review of undergraduate programs, health care professions programs, high-cost engineering programs, and "excess academic capacity in southeastern Michigan." It also calls for adherence to capital outlay criteria developed by the Commission as well as specified criteria for closing or merging institutions.

With regard to undergraduate programs, the Commission recommends specifically that the Department of Management and Budget and the Department of Education review "non-core" programs that are "low degree producing" as a part of the annual budget review process that "the Executive Office and the Legislature be advised to discontinue subsidies for those programs." The Commission also recommends that the Department of Management and Budget undertake a detailed analysis of costs and benefits related to academic programs in southeastern Michigan for the purpose of determining whether some of these programs should be consolidated.

Conclusions

What are some of the conclusions that might be drawn from these recommendations?

First, the Commission clearly wants state policymakers to make more of the decisions about institutional roles, educational programs, and methods of funding than they have done in the past. This is reflected in numerous statements throughout the final report, including some of the quotations presented above.

Second, the Commission has apparently assumed that some of the institutional governing boards would be willing to accept limitations on roles and missions of their respective institutions if state policymakers can agree on a classification system of some kind. The Commission makes no reference to the constitutional responsibilities of these governing boards or the need for obtaining their support. Instead, it leaves the impression that these responsibilities can be ignored if decisions at the state level are somehow tied to funding of educational programs.

Third, the Commission seems to have ignored some of the problems involved in making decisions at the state level about funding of educational programs as opposed to funding of institutions. Educational programs represent combinations of courses, many of which are offered by organizational units other than those identified with the programs. As a result, and particularly in a period of retrenchment, the marginal costs of programs may not be nearly as large as might be expected. Furthermore, when these costs are weighed against the intellectual and professional benefits to students and faculty, something that can only be done within the institutions, the costs are even less important. Yet the Commission seems unaware of such complexities.

Fourth, the Commission has taken the position that the state budgetary process will continue to provide the best mechanism for planning and cooridnation of higher education in Michigan if it can be strengthened by cooperative efforts involving the Governor and the Legislature, a position that will likely be supported by institutional representatives. However, the Commission has also indicated that the Department of Management and Budget and the Department of Education should play more important roles in this process than they have done in the past. Most of the leadership at the staff level in recent years had come from the Senate Fiscal Agency on some matters and the House Fiscal Agency on others. It appears, therefore, that the Commission has attempted to shift that leadership to the Department of Management and Budget and the Department of Education.

Fifth, the Commission leaves the impression that employees of the Department of Education, including the State superintendent of Public Instruction, can act independently of the State Board of Education. Yet the Department is legally responsible to the Board. And the State Superintendent of Public Instruction is appointed by and serves at the pleasure of the Board. It would seem, therefore, that the Commission has attempted to strengthen the role of the State Board of Education in the state budgetary process without making direct references to it in the various recommendations.

Sixth, and last, the Commission leaves the impression that the only alternative to its recommendations is a "super board" of some kind. Persons familiar with the literature on planning and coordiantion of higher education might find it difficult to accept the idea that a separate state board for higher education with advisory powers only would be classified in this manner. Such a board was recommended by a commission appointed by the former governor in the early 1970's. That board would have had less authority over public instituions of higher education than the State Board of Education now has. The recommendation, however, would have required a constitutional amendment.

Prospects for the future.

These and other matters related to the final report of the Governor's Commission on the Future of Higher Education in Michigan are now being considered by a Senate Select Committee on Higher Education established for this purpose. Hearings are being held so that institutional representatives and others will have opportunities to present their views on the various recommendations.

The future of decentralized control for public institutions of higher education in Michigan will clearly be influenced by these deliberations. It is too soon to know, of course, whether members of the Legislature will support proposed restrictions in institutional activities.

The continuing challenge, it would seem, is for the Governor, the Legislature, and institutional governing boards to work together to maintain that delicate balance between institutional goals and statewide goals that is so important in terms of decentralized control.

NOTE: This paper was presented to the Senate Select Committee on Higher Education, Lansing, Michigan, March 5, 1985.

The author is a Visiting Scholar, Center for the Study of Higher Education, the University of Michigan, and Professor Emeritus, Department of Educational Administration, Michigan State University.

AN OVERVIEW OF INSTITUTIONAL EVALUATION*

Richard Miller

I would like to begin with a perspective on institutional evaluation that fits in with a reoccurring theme that has run through this truly excellent conference--one that conference participants, coming from quite different points-of-view, have articulated: It is the importance of students. Clark Kerr said in 3000 Futures that this decade may be "the golden age for students," and Alexander Astin's book has been mentioned. At least three of the state coordinating board reports focused on student outcomes; Charles McClain said that the present dialogue on assessment/evaluation has focused the agenda on student learning; and Allan Ostar emphasized that students must come first.

The perspective that I would like to add comes from a small classic piece by E. R. Guthrie in 1954:

> Students do not enter college with a thirst for those virtues that make middle age and old age interesting. Hot-rodders and drum majorettes do not become scientists or philosophers, but they do enter college and do become voting citizens. They also build our factories, fight our battles, and make our laws. It is their country."

Institutional evaluation very likely dates back to the early days of formal, Western education, and one can imagine that a robust evaluative dialogue must have taken place in ancient Greece between the merits of Socrates and his school of orators, and Aristotle and his academy.

Institutional evaluation in the United States can be dated back to the early days of Harvard, or even to September 3, 1642. On that day, John Winthrop, Governor of the Massachusetts Bay Company, journeyed from Boston to Cambridge and presided over the public final examination given to nine seniors of Harvard College by its external Board of Overseers (Harclerod, 1980, p. 1).

Members of the higher education sector currently have ample opportunity to make a decisive impact upon the growing trend toward program and/or institutional evaluation if we give sufficient attention to meaningful research, if we exercise political savvy, and if we oppose creatively yet vigorously those who propose simplistic and misleading solutions to complex problems. In many cases it is counterproductive for us to propose further study without action expectations. Sometimes the external demands for decisions will not wait very long. Some of these points were articulated also last evening by Hoke Smith.

In addition to the confirmation of increased institutional evaluation that is in evidence in our presentations here, there are other evidences. For example, the July 2, 1986 issue of The Chronicle of Higher Education carried the initial findings of a very recent A. C. E. survey under the heading of "Assessment of Learning Backed, but Misuse Feared, Survey Finds." The chief academic officer survey found that 91 percent said that "assessment should be linked to instructional improvement"; 47 percent said that they feared "misuse," and 43 percent said that "most faculty do not support it"--it presumably being more evaluation.

General Directions in Institutional Evaluation

A number of general directions in institutional evaluation may be identified. I have listed five.

1. More institutional evaluations will be undertaken, both from external pressures such as business/industry, and from state and federal governmental forces.

The state legislatures in their 1985-1987 appropriations for higher education have been more generous than in the previous several years, due largely to the current emphasis on quality education, to reasonably good state level and national economies, and to feelings on the part of state officials that higher education has been squeezed enough--at least for awhile.

But there is another side also, that of increasing restiveness on the part of external forces about accountability, or more precisely, our alleged lack of it among colleges and universities. Our colleges and universities may be managed better than the external forces believe is the case, but even with this judgment there seems to be a malaise about effective and efficient management among those in academe who do it for a living. This serendipity may be short lived due to predictions of declines in student enrollments as well as in the economy. With the need for more money for operations and fewer students in the classrooms, the external pressures for accountability may become greater.

And the trend toward greater accountability does not stop at our shores, as those of you know who have been following the recently formed National Council on Evaluation in France, or the unprecedented year-long examinations of Britains 44 universities by special committees established by the University Grants Committee. Every department in these institutions were evaluated on a five-point scale, primarily on these four criteria: (1) each department's five best research publications as chosen by the professors in that department; (2) the opinion of research councils and distinguished academics: (3) the amount of money attracted from industry; and (4) the number of professors who have received distinctions such as Fellowship in the Royal Society. The ratings were announced in letters sent to each University in mid-May.

A second general direction has been touched upon several times thusfar in this conference: The United States is experiencing an unprecedented surge of interest in quality education.

Never before in our history has so much attention been concentrated on quality education in reports, television programs, articles, conferences and conversations.

This circumstance augurs well for improving all facets of education provided educators take advantage of this ground-swell of interest by working more diligently for greater improvements in the quality of education. As yet, we see only the beginnings of qualitative improvements but we hope that the trends will continue in a positive manner. It is difficult to understand the plethora of national reports that have been issued to improve our system of higher education, but it is obvious that the current surge of reports and interests in quality education is greater than any previous one. If only 20 percent of the most prominent recommendations are implemented, which is about the efficiency of a gasoline engine, we should be quite pleased.

3. The rapid growth in interest in value-added education and in student outcomes measurement will continue for a number of years.

The value-added approach, simply stated, entails measurement at the beginning of a course and/or program and another measurement at its conclusion. The difference is what we call "value-added" because of the program. The term "student outcomes" refers to what knowledge, values, or attitudes that the student has mastered at its conclusion, as measured by some kind of test. Most often the outcomes are knowledge or factual information rather than values and attitudes.

Much discussion and literature in recent years has focused on student outcomes; in other words, what does the undergraduate student take from a college after four years? In a recent article, Peter Ewell wrote: "What should students be expected to know by graduation? Much of the difficulty in establishing assessment programs is the general lack of agreement on the answer to this question" (1985, p. 35).

Some earlier, simplistic approaches have given way to more sophisticated ones. Student outcomes have been measured in many fields in a number of ways, including tests for professional certification or licensure. A wide array of these tests cn be found in Burros's Mental Measurements Yearbooks, and the latest edition lists standardized instruments for over 40 subject areas.

The current extensive program revisions in the general education component in our colleges and universities is evidence that testing for student outcomes should include greater attention to the broad aesthetic and cultural aspects as well as professional preparation. The American College Testing Program has developed the College Outcome Measurements Project (COMP) for assessing general education knowledge and skills. The outcomes test covers the six themes of communicating, solving problems, clarifying values, functioning within social institutions, using science and technology, and using the arts. The outcomes assessments at the University of Tennessee, Knoxville and at Northeast Missouri State University are making extensive use of COMP test results.

Chambers has written that for outcome measures to be used for evaluating educational quality they should meet two tests: "First, each outcome should represent something that is teachable; that is, it is generally accepted that the particular educational objective adopted can be achieved through an instructional or didactic process. . . . Second, each outcome should measure something judged relevant to the graduate's functioning throughout adult life" (1979, p. 31).

Some individuals have confused the value-added approach and academic standards, and somehow have alleged that the value-added approach will lower academic standards. One can make the case, quite to the contrary, as Alexander Astin has done, tht considering academic standards as the same as value-added standards can "provide a much more vigorous and far less ambiguous set of standards than current academic standards based solely on relativistic course grades" (1983, p. 136).

4. *The rating game will continue*.

Every year numerous ratings of colleges and universities provide information on the best party schools, campuses with the most active sports programs, those with the prettiest or most handsome students, most beautiful campuses--and other equally significant information on our academic institutions.

Perhaps symbolic of the current popularity of campus ratings in the cover story article in the November/December 1984 issue of *Change* magazine, which was: "Who Is Jack Gourman: And Why Is He Saying All Those Things About My College?" (The first Gourman report was published in 1967 with little fanfare and even less notoriety.) And the cover story article in the November/December 1985 issue of *Change* magazine was titled: "Are Colleges Making the Grade?"

To learn more about how students select colleges, The Carnegie Foundation for the Advancement of Teaching surveyed 1,000 high school seniors about the sources of information that they used most frequently in deciding on a college. The survey "revealed that nearly 60 percent of all prospective college students read commercial college guides. The students and their parents rated the relevancy and accuracy of these guides higher than college publications. But trying to pick a good guide is almost as complicated as trying to pick a college" (1986, pp. 30-31).

The ratings can provide useful information, however. They can assist parents and counselors in sorting through our over-3000 post-secondary institutions; they can assist graduate students in making educational decisions; they can provide institutional yardsticks; and they can provide governing boards with comparative information. But they also can provide simplistic and sometimes distorted information about complex problems.

5. *We are becoming more proficient in undertaking institutional evaluations*.

Our improvement over the past 10 years has been due to considerable practice in this craft. While practice does not make perfect, we have learned from our mistakes. Also, we have begun to use more effectively the technologies at our disposal. The work at the National Center for Higher Education Management Systems (NCHEMS) has provided earlier leadership in developing evaluation systems.

What Kind of Institutional Evaluation?

What kind of evaluation can be most useful? What Kipling once wrote--"There are nine and sixty ways of constructing tribal lays, and-every-single-one-of-them-is-right"--points up the importance of developing an approach to evaluation tht is tailored to the specific institution. Yesterday, Charles McClain used South Dakota's adoption of his institution's model as a very questionable practice.

Two general types of evaluation can be identified. One type consists of those evaluations that are less than institution-wide in scope, such as the assessments of academic programs, faculty, students, and other areas. However, a systematic and extensive plan for academic program evaluation may become more closely akin to the second type, which is institution-wide evaluation.

The second type of appraisal, that of the entire college or university, has received less attention, although interest in this dimension now is growing substantially in many countries. One can identify at least five approaches to institutional evaluation. These are 1. educational auditing, 2. external consultant reviews, 3. self-studies for accreditation, 4. self-studies for institutional purposes, and 5. state and federal reviews. A brief discussion of these five approaches is included as an appendix.

General Guidelines for Institutional Evaluations

If you are considering or planning to undertake a major study of institutional quality, what are some guidelines that you may want to consider? Eight guidelines are suggested for your consideration:

1. Theoretical and operational plans for evaluating institutional quality should be compatible with the organizational ethos.

Simply put, one should abide by the classic anthropological concept of "Adapt not adopt." The argument for "adoption" is that the internal structure and dynamism of a plan is less likely to be destroyed when the whole concept or model is adopted.

On the other hand, each college or university has a unique history and nature that can result in forcing a square into a round hole if pure adoption is pursued. Our experience leans strongly toward modifying and tailoring concepts and plans to fit the circumstances and needs of each situation. Or, as Allan Ostar commented: "Make the plan "institution-specific."

1. Vigorous yet sensitive administrative leadership is essential.

This leadership includes providing the initial impetus for the evaluation, keeping track of progress and/or problems, and seeing that the emphasis is toward some ultimate concrete recommendations for action.

The manangement of higher education is receiving renewed attention as the skills for managing institutional decline have become more prevalent in recent years than those needed for managing institutional growth. And the matter is made more complicated because today's senior administrative officers have a three ring circus to contend with. In ring one is academic programs that have declining student enrollments, in ring two are those with stable enrollments, and in ring three are programs that are or can be increased in student size. And such adjustments often are made in an overall budget that is only modestly increasing, or in some cases it is decreasing.

But what kind of leadership are we talking about? A debate about the most effective styles of academic leadership has been going on for some time. The 1974 book by Cohen and March, titled, Ambiguity and Leadership, received wide attention in the U.S.A. Its findings were based upon data gathered primarily during the student unrest in the late 1960's. They described the university as "a prototype organized anarchy" (pp. 2-3).

Their thesis did, and does, strike a responsive chord for the countless frustrations, uncertainties, and uncontrollables in higher education administration, yet whenever a nation, a family, or a university comes upon hard times, more vigorous leadership is called for. The extensive academic cutbacks in most U.S. colleges and universities in the past 10 years have brought forth the need for more dynamic leadership, which is quite contrary to the sense of presidential helplessness developed by Cohen and March.

Writing on "Effective Administrators", David Whetten categorized administrators as either gatherers or hunters. Gatherers forage passively in an immediate area for enough food to satisfy daily needs. Hunters aggressively pursue a quarry as far as necessary and recognize that survival depends upon cunning and strategy. According to Whetten's research, the aggressive leadership style of many academic administrators possesses three distinguishing characteristics:

> First, assertive academic administrators preserve and highlight sources of opportunity within their institutions . . . . A second component of assertive leadership is cultivating a distinctive institutional image. The third component of assertive leadership touches on the role of charismatic leadership, especially in the strategic planning process. Effective administrators recognize that in charting a course for an institution, inspiring vision precedes, rather than is driven by, quantitative analysis (1984, pp. 41-42).

I was very pleased to see that June 4th Chronicle article about the forthcoming book on Searching for Academic Excellence by Wade Gilley, Ken Fulmer, and Sally Reithlingshoefer--which emphasizes the positive. We have heard much about how bad things are, and the Carnegie report on the status of faculty released last fall must be placed in this category. It is good to know that we are not "all going to hell in a handbasket."

In sum, the climate and circumstances in the U.S.A. now favor the kind of vigorous leadership that is a critical factor in improving institutional quality. There are some positive signs that the chief executive officer styles are reflecting this need more now than in the past two decades, but the mills of the Gods grind slowly . . . .

A third general guideline: An overall plan for enhancing institutional quality should be developed and communicated.

Members of the academic community should have ample opportunity to know that an evaluation effort is underway, with certain individuals having some designated responsibilities, with mandates and/or charges established to guide the effort, with some reporting dates scheduled, with open hearings planned, and with action recommendations anticipated. Thomas Hegarty said yesterday, based upon his experience with the Potsdam APEP project, said that "information-sharing is perhaps the most important aspect of communication."

Characteristics of institutional evaluation have been developed by a number of groups and individuals. The Northwest Association of Schools and Colleges, for example, determined that the self-study process should be flexible and permit different approaches; be attuned to current institutional priorities; utilize recent institutional research or studies; involve as many people as possible; have sufficient breadth and depth to review accountability; focus on results of the educational program; use information and results to improve the document; yield a concise, readable useful document; and foster ongoing self-study and planning (1975, p. 6).

The New York State Education Department project on self-assessment for colleges and universities suggested the following steps in self-assessment:

1. A clear definition of the goals of the assessment, as distinct from the goals of the specific area being assessed, should be made. The main focus of assessment should be on evlaution of educational quality as measured by goal-oriented outcomes.
2. All persons who are affected and interested in the programs under review should be continually made aware of and often involved in the assessment process. Responsibility for setting priorities, designing the assessment, collecting and analyzing data, and evaluating and using them should be assigned to appropriately skilled persons.
3. A determination of how well the goals of the specific area assessed are being met should be made. The appropriate instruments and techniques must be selected and administered to the constituencies involved (for example, administrators, faculty, students, graduates, employers, and outside groups).
4. The process of collecting data should be established in such a way that it can continue beyond the first self-assessment as a routine function of the master planning and decision-making process.
5. Analysis of data, reporting of findings, and recommendations for action should be carefully monitored by the person(s) responsible for the self-assessment. Priodic follow-up of recommendations is essential to determine if any actual results have occurred.
6. Essential to effective self-assessment is the periodic evaluation of the system itself. The system should be cost-effective in both dollars and human time spent to provide vital information for decision making (1979, pp. 5-6).

4. Models and plans for studying and evaluating institutional quality should be linked to resource allocation and/or re-allocation.

It has been said that ignorance is a great innovator because we never would have undertaken many of our accomplishments if we knew what we were getting into when we started!

Accepting the earthly wisdom of the previous statement does not negate the reality of a French saying, which is translated as: "Power is where the money is, and the rest is theatre." One does not have to agree fully with the statement to appreciate its pertinence to academic institutions.

The importance of carefully linking academic plans/models to human and material resources has been forced upon most of us by having to cope with diminishing resources. Still, in the U.S.A. we have a tendency to loosely couple, if we couple at all, academic models and finances. Most collegiate administrators believe they do relate models and plans to finances but upon closer scrutiny this relationship is more often toward the end rather than initially and throughout the process.

5. Institutional evaluation should use objective data where available and purposeful but make no apologies for using subjective data.

Or, it is better to be generally right than precisely wrong (Enthoven, 1970). Objective data is important, yet considerable variation exists in its availability and quality. The absence of objective data should stimulate those responsible for institutional evaluation to devise their own survey instruments, guidelines, and checklists or to use systematically treated judgement as bases for decision making. The lack of "hard" data should not deter careful and systematic decision making about important institutional matters. Solid bases for decision making can be developed by using whatever "hard" data are available along with experience, judgement, and common sense. Important institutional decision sometimes are made on much less.

6. Those developing models and plans for evaluating institutional quality should have an error-expectation attitude.

With this mind-set it is less likely that one will become discomposed and discouraged when things go awry. It may help if one takes to heart the earthy wisdom found in two of Murphy's laws:

- Nothing is ever as simple as it seems.
- If anything can go wrong, it will.

An accompanying important mind-set is that of retaining the mental dexterity necessary for responding to unexpected opportunities. One of the most dramatic examples of this mental dexterity is the quarterback in American football. Competent ones have the ability to "check off" at the line of scrimmage; that is, they can call an audible play which means changing the upcoming play on-the-spot based upon hw he "reads" the defensive formation that is opposing him. This mental dexterity is related to a point that Charles McClain made yesterday when he said that the future will unfold in a meandering way; move on a fuzzy vision; and do something, letting others fine-tune.

How many potentially excellent models and plans can you recall that were fatally or near-fatally afflicted with the rigor mortis syndrome, that is, we have an outstanding plan that we must follow because we must follow it; or with the blinders syndrome, that is, one can see very well straight ahead but is oblivous to everything that is going elsewhere?

7. Institutional evaluation studies should be action-oriented, with plans for moving reports-to-action.

"Filed and forgotten" should not be ascribed to most institutional evaluation efforts, considering the investment of thousands of hours and dollars and the importance of keeping the institution abreast or ahead of its problems, but the history of institutional studies does not provide cause for exuberance about effective implementation of recommendations. Today, however, there are a number of excellent case studies of institutional evaluations that were action-oriented because the times demanded more rigorous approaches to implementing recommendations.

The administrative charge to an institutional evaluation committee should make clear that its work plan is geared toward recommendations-for-action. This position from the outset will remind those involved in the projects, and others, that their efforts are less likely to be filed and forgotten.

8. A plan for evaluating the evaluation should be included.

Most evaluation reports make little or no provision for evaluating their effectiveness. Such evaluations of evaluations can be useful as testimony to the importance of evaluation in future improvement, for providing systematic checkpoints of progress, and for providing a procedure for orderly modifications based upon subsequent findings. Thomas Redmon, yesterday in his six steps in making policy analyses, recommended the sixth step as evaluation of what you have done, not so much in terms of present details but in terms of possible future consequences and impact.

Some Problems in Evaluating Institutional Quality

There are many potential problems involved when one begins a serious effort toward evaluating institutional quality. And I am certain that you can add several problems that you have encoutered:

1. Poorly defined goals and objectives can significantly impede the effort.

If we do not know where we are going then any road will take us there.

Statements of goals and objectives may be flawed in a number of ways; some of the common shortcomings are:

A. Projects goals and objectives are not related to overall university goals and objectives.
B. They are not closely related to the major tasks to be accomplished.
C. They are too low to challenge, or they are too utopian to achieve.
D. The focus is on activities rather than upon outputs.

E. They are too wordy and full of "academic garbage," and/or they address unimportant matters.
F. They are written more to impress then to implement.

2. Failure to give adequate attention to the processes involved in bringing about improvements can be fatal to the enterprise.

In the U.S.A. we are beginning another cycle of interest in the processes of change. A previous cycle was sparked by the Kennedy-Johnson era in the 1960's when much literature, research, and conversation took place about the "how" dimension. We are moving into this area of interest again, motivated primarily not by growth but by steady-state or decline.

There are differences in the meanings of change that are deeply rooted in academe. At the risk of oversimplification, it is a perceptual problem rooted in differences between academic disciplinarians and administrators. The latter group focuses more on the "bottom line," or on what is produced, while the academicians focus more on the "how" dimension, or the processes involved.

The management of change in U.S.A. colleges and universities has some genuine problems, and these flaws become more apparent when the topic is as complex, important, and sensitive as evaluating institutional quality. The title of one chapter in the book by George Keller, "Slouching Toward Strategy," captures the problem. Also an earlier quotation from Cohen and March's book that spoke of the university as "organized anarchy". These characterizations apply even more to the way academic committees go about their work than to the academic adminsitration of the institution.

Typical academic committees can be defined as gatherings, usually casual, covened late, without agendas, that take minutes but waste hours, although often they do not even take minutes. One faculty member said that when he died he hoped that it would be in a faculty committee meeting because there the transition from life to death would be scarcely perceptible!

Yet I agree that faculty involvement must be open, persistent, meaningful, and appropriate. The theme of "faculty ownership" was mentioned several times yesterday by administrators in talking about faculty members. But as a faculty member who remains slightly tainted as an area coordinator, I would suggest that we may tend to build up the virtues of faculty members when away from home at conferences, and we can overlook the fact that tenured faculty members can wait-out and wear-out administrators through their intricate knowledge of the mystical labryinths of academic decision-making or the lack thereof.

3. A "cargo-cult" mentality can cause problems.

Many years ago the natives in the South Seas islands believed that a ship laden with gold and precious gems would some day sail into their harbor and bring a happy ending to all their problems.

Some of our approaches to institutional evaluation bear a resemblance in that we latch upon a single answer to complex problems. The latest example is the tendency on part of some American educators to look to the measurement of student outcomes as the answer to evaluating institutional quality. The overly simplistic answer was raised last evening by Hoke Smith.

In the mid-2970's some individuals looked to fiscal indicators as the answer.

How nice it would be if institutions could be evaluated so simply. Evaluating institutional quality, if it is anything, is a many splendored thing that requires several component parts that form a whole, which we hope will be greater than the sum of the parts.

Retuning for a moment to the outcomes or value-added approach: Care must be taken that we do not ascribe values to outcomes test scores that are not intended. On this point, Peter Ewell notes: "Interpreting every score gain that we observe on a test as an actual change in student learning can be a very risky activity (1985, p. 18). And the title of Herbert Spencer's little book in 1848 is quite relevant. It asked: "What Knowledge Is of Most Worth?"

4. Making too many changes, too fast can be a problem.

Generally speaking, academe is not known for its rapid changes. It has been said that if Ford Motors were a university they would still have a department for the Edsel.

Extensive and complicated changes take time. In his article describing the exemplary value-added program developed at Northeast Missouri State University, Osigweh traces steps in this program back to 1973 (1986, pp. 28-33).

In a world where instant coffee, fast diets, quick divorces, and instantaneous worldwide communications are commonplace, is it any wonder that we might become impatient with multi-year academic improvement efforts? Perhaps we all need to remind ourselves occasionally that crash programs usually crash.

5. Failure to anticipate problems can provide short-term euphoria and long-term extinction.

While "ignorance can be a great innovator," the other side of the coin is that failure to ameliorate weaknesses and problems can be of great advantage to the opposition. Successful athletic coaches spend much time on strengthening weaknesses and problem areas to avoid such an eventuality.

In order to better anticipate problems, it might be useful to consider several reasons why evaluation models and plans fail:

A. An obvious reason, but one that we prefer to overlook is that they are poor plans. Our projects and plans in higher education rarely fail. Crop failure analyses are common in agriculture and autopsies are common in medicine, but in higher education we just write another proposal.

B. The model or plan is not vigorously and consistently supported at the top. A 1974 study by Davis and Batchelor found that only seven percent of college presidents really got fully involved in the planning process (1974, pp. 32-36). Since the study is several years old, one would like to thtink that some improvement has taken place in the intervening years.

C. Failure of evaluation programs is possible if the enterprise is viewed as the preserve of the institutional planners. Institutional researchers and planners are critical, but their roles are "means" rather than "ends." It is interesting to note what is happening in U.S.A. industry. The September 17, 1984 issue of *Business Week* magazine titled its front cover and feature story, "The New Breed of Strategic Planner." Many major corporations have cut their planning staffs to the bone, and middle management personnel and some line personnel have been given planning "know-how" and planning responsibilities. This rather dramatic shift is consistent with the growing emphasis on the entrepreneurial spirit in American business and industry that is developed in Rosabeth Kanter's provocative recent book on *The Change Masters*.

D. Plans for developing institutional quality are failure-prone if they do not carefully consider the iceberg-under-the-surface, the *detailed managerial system*, that is needed to develop, maintain, evaluate, and modify institutional evaluation plans.

E. Another formula for failure is the misplaced use of evaluation as a smoke screen. A few administrators have sought to keep people busy planning and operating so they cannot turn to the real problems, one of which may be the administrator!

F. Evaluation plans can be flawed if they do not have a built-in and identified "fixer". This is the English translation for a Russian individual that is prominent in the extensively planned Soviet sytem--the "tolkash." A procurement specialist, this individual is openly known but officially a non-person. He or she adds fluidity, creativity, and success to an often awkward planning system.

A "fixer" in our collegiate planning process may be a senior faculty member, may be someone in the administrative offices, or elsewhere, who is able to add fluidity, creativity, procedural adjustments, and evaluations to our campus planning systems. By leaving nebulous the responsibility for adjustments in the planning processes and systems, we risk significantly weakening their impact.

G. A seventh possible cause for failure is lack of attention to a positive and sustained mind-set. Every middle and senior level collegiate official might replace the IBM slogan of "Think," with "Think Strategically." On this point, King and Cleland wrote: "The ability to 'think strategically' in comprehensive policy terms is perhpas the least homogeneously distributed personal managerial trait" (p. v.).

Finally, a perspective: These thoughts about improving institutional quality need to be kept in the larger context of a college education. I believe that it must remain more than an accumulation of credit hours or a pipeline to a position as important as are these accomplishments. There is a spirit about a campus that sets it apart as an unique interlude. Its products and its on-going activities account for most of human progress and inventions; its stored and alive sources of knowledge contain and carry forward the magnificent heritage and knowledge of mankind for all to ponder, to question, and to know. Its traditions and mores cause frustrations and misunderstandings among its critics; its freedoms provide an essential bulwark for the democratic way of life; and its concern for idealism and ideas has kindled the hopes and dreams of millions who have found themselves during college days. Most of us are better people because of this brief period, and this is what counts.

Richard I. Miller is Professor of Higher Education at Ohio University, Athens, Ohio.

References

American College Testing Program. The Institutional Self-Study Service. Iowa City, Iowa: American College Testing Program, 1970.

Astin, A. W. "Strengthening Transfer Programs." In Vaughan, G. B. (ed. , Issues for Community College Leaders in a New Era. San Francisco: Jossey-Bass Publishers, 1983.

Bowen, H. R. "Goals, Outcomes, and Academic Evaluation." In Evaluating Educational Quality: A Conference Summary. Washington, D.C.: Council of Postsecondary Accreditation, 1979.

Carnegie Foundation Survey of the Transition from High school to College. "How Do Students Choose A College." Change. 1986. 18. 29-32.

Chambers, C. M. "What have We Learned?" In Evaluating Educational Quality: A Conference summary. Washington, D.C.: Council on Postsecondary Accreditation, 1979.

Cohen, M. D. and March, J. G. Leadership and Ambiguity: The American College President. New York: McGraw-Hill. 1974.

Council on Postsecondary Accreditation. The Balance Wheel for Accreditation. Washington, D.C.: Council on Postsecondary Accreditation, 1976.

Davis, J. S.. and Barchelor. S. A. The Effective College and University Board: A Report of a National Survey of Trustees and Presidents. Research Triangle Park, N.C.: Research and Evaluation, 1974.

Enthovan, A. "Measures of the Outputs of Higher Education: Some Practical Suggestions for Their Development and Use." In B. Lawrence and associates (eds.), Outputs of Higher Education: Their Identification, Measurement, and Evaluation. Boulder, CO: Western Interstate Commission for Higher Education, 1970.

Ewell, P. T. "Academe Nervous as Assessment Acquires New Political Meaning: Testing of Students to Evaluate Both Them and Their Institutions." The Chronicle of Higher Education, November 27, 1985.

Ewell, P. T. "Assessment: What's It All About?" Change, 1985, 17, 32-36.

Gourman, J. The Gourman Report: Ratings of American Colleges. Phoenix, AZ: The Continuing Education Institute, Inc., 1967.

Harcleroad, F. F. "The Context of Academic Program Evaluation." In E. C. Craven (ed.) Academic Program Evaluation. New Directions for Institutional Research, Number 27. San Francisco: Jossey-Bass, 1980.

Harcleroad, F. F., Educational Auditing and Accountability. Washington, D.C.: Council on Postsecondary Accreditation, 1976.

Harcleroad, F. F., and Dickey, F. G. Educational Auditing and Voluntary Institutional Accrediting. ERIC/Higher Education Research Report, No. 1. Washington, D.C.: American Association for Higher Education, 1975.

Kanter, R. M., "Heralding a Renaissance for Corporate America--Relying on the People." (Excerpted from Kanter's book on The Change Masters) Change, 13:18+, July/August, 1984.

King, R. W. and Cleland, D. I. Strategic Planning and Policy. New York: Van Nostrand Reinhold Company, 1978.

Keller, G. Academic Strategy. Baltimore: The Johns Hopkins University Press, 1983.

Ladd, D. R. Change in Educational Policy: Self-Studies in Selected Colleges and Universities. New York: McGraw-Hill, 1970.

Manning, B. A. The "Trouble Shooting" Checklist for Higher Educational Settings. Austin: University of Texas Research and Development Center for Teacher Education, 1976.

Marchese, T. "Let's Reward Quality: Tennessee's Bold Experiment." Change, 1985, 17, 37-47.

Mautz, R. D., and Sharaf, H. A. The Philosophy of Auditing. Chicago: American Accounting Association. 1961.

Mayhew. L. B., and Ford. P. J. Changing the Curriculum. San Francisco: Jossey-Bass, 1971.

Miller, R. I. (ed.) Institutional Assessment for Self-Improvement. New Directions for Institutional Research, Number 29. San Francisco: Jossey-Bass, 1981.

Miller, R. I. "Appraising Institutional Performance." In Jedamus, P., M. W. Peterson and associates. Improving Academic Management, Jossey-Bass, 1980.

Miller, R. I. The Assessment of College Performance: A Handbook of Techniques and Measures for Institutional Self-Evaluation. San Francisco: Jossey-Bass, 1979.

Newman, F. Report on Higher Education. Washington, D.C.: U.S. Department of Health, Education and Welfare, 1971.

New York State Department of Education. A Handbook for Self-Assessment for Colleges and Universities. Albany: New York State Department of Education, 1979.

Northwest Association of schools and Colleges, Commission on Colleges. Accreditation Procedural Guide. Seattle: Northwest Association of Schools and Colleges, 1975.

O'Connell, W., Jr., and Meeth, L. R. Evaluating Teaching Improvement Programs. New Rochelle, N.Y.: National Teaching Program, 1978.

Osigweh, C. A. B. "A 'Value-Added' Model of Measuring Performance." College Teaching. 1986, 34, 28-33.

Porter, D. "Developing Performance Indicators for the Teaching Function". Paper presented to the CECD/CERI-IMHE Program fifth special topic workshop. Paris, France, June 5-8, 1978.

Pottinger, P. E. "Some Comments on the Effective Use of Consultant and External Advisors." Paper presented at project directors' meeting of the Fund for the Improvement of Postsecondary Education. Boston: Institute for Competance Assessment, Division of McBer and Company, 1975.

"Report of the Visiting Committee to Evaluate the Department of Anatomy." Universitv of Chicago Record, 1975, 3, 109.

Sizer, J. "European Perspectives Suggest Other Criteria." In Miller, R. I., Institutional Assessment for Self-Improvement. New Directons for Institutional Research, Number 29. San Francisco: Jossey-Bass, 1981.

"The New Breed of Strategic Planner" (cover story) Business Week, September 17, 1984.

Whetten, D. A., "Effective Administrators." Change, 1984, 16, 38-43.

APPENDIX*

This section will consider these five different approaches to institutional evaluation: 1. educational auditing, 2. external consultant reviews, 3. self studies for accreditation, 4. self studies for institutional purposes, and 5. state and federal reviews.

1. Educational Auditing. The Securities and Exchange Commission was established in 1934 in response to serious questions about the conduct and disclosure of business operations. In 1917, the American Institute of Accountants had published a list of standards for the preparation of audits. Although the form and nature of the audits vary, the following eight basic postulates were developed by the American Accounting Association (Mautz and Sharaf, 1961, p. 42):

1. Financial statements and financial data are verifiable.
2. There is no necessary conflict of interest between the auditor and the management of the enterprise under audit.
3. The financial statements and other information submitted for verification are free from collusive and other unusual irregularities.
4. The existence of a satisfactory system of internal control eliminates the probability of irregularities.
5. Consistent application of generally accepted principles of accounting results in the fair presentation of financial position and the results of operations.
6. In the absence of clear evidence to the contrary, what has held true in the past for the enterprise under examination will hold true in the future.
7. When examining financial data for the purpose thereon, the auditor acts exclusively in the capacity of an auditor.
8. The professional status of the independent auditor imposes commensurate professional obligations.

John Sizer (1981) asks whether five standards used by the American Accounting Association can be applied to higher education institutions. These five standards are: relevance, verifiability, freedom from bias, quantifiability, and economic feasibility (pp. 75-79). And Porter (1978) proposed that the test of institutional acceptance--are the indicators fair and acceptable--be added to the five standards.

As pointed out by Harcleroad and Dickey (1975, p. 15): "With some slight modification and changes in terminology, these same postulates might very well be applied to the process of institutional auditing and accrediting. The verfiability of date, determination of adequate sampling, and probability theory would be characteristic of both business and educational institutions."

---

[*] This material is based upon cited reference by the author in the bibliography plus lecture notes.

Of the various differences that exist between the business and educational audits, perhaps the most important one relates to scope. The fiscal audit is only a part of the educational audit, but it is the significant aspect for business. Some individuals, particularly those with business backgrounds, may relate institutional health almost exclusively to fiscal health. Such close equating should be avoided by institutions of higher education because of fundamental differences in puposes and funding procedures.

Harcleroad (1976, p. 18) writes that educational accreditation can profit in a number of ways from some of the experiences of the auditing profession:

> (1) There might be improvement in the total process if more of the work were to be carried on by full-time professionals, without losing the value of having many of the participants serving on a part-time basis as at present. (2) There is a need for a research staff, similar to that of the Financial Accounting Standard Board, to be working on basic principles of assissment and procedure to followed. (3) Some form of continuity is needed for the institution involved and the main members of the team that will work with it as a representative of the accreditation body. (4) The standards of the regional associations and the specialized associations, and their respective commissions, should be similar enough to discount claims that great differences invalidate the entire process. The business auditing system, even with all its current problems and its many auditing firms, has sufficient comparability in its standards and their application to be quite credible, most of the time.

Private managerial firms are becoming interested in the audit approach to higher education; and the first report by Newman (1971, p. 70) spoke of the creation of "new regional examining universities." In considering the need for evaluating higher education, Bowen (1979, pp. 1-21) wrote that one of the needs of higher education is to find a means of evaluation that is genuinely disinterested and yet takes account of the many intangible elements. Perhaps a new profession of independent judges of productivity and performance should be created to evaluate institutions as well as higher education as a whole.

2. External Consultant Studies. A number of institutions of higher education as well as state systems have chosen to use external consultants for departmental and institutional studies. The New York Board of Regents study of doctoral program quality, the Illinois Board of Higher Education's Commission of Scholars, and the Louisiana Board of Regents Doctoral Program study all relied upon external consultants as a primary data source. The University of Chicago, for example, makes extensive use of recognized scholars from similar universities in analyzing its various programs. Their reports are printed in an official publication of the university called The University of Chicago Record.

The use of external reviewers/consultants carries the primary advantages of expertise and impartiality. One disadvantage is characterized in the "Report of the Visiting Committee to Evaluate the Department of Anatomy" (1975, p. 109) of the University of Chicago: "None of us is confident that a single experience, however concentrated, can reveal all the details of a complete system, nor that our responses are the most appropriate ones possible." The use of the phrase "the most appropriate ones" implies another possible disadvantage: institutional repercussions from ill-advised remarks, statements, or conclusions. The reviewers leave on the afternoon airplane, but those remaining must live with the report. The reviewers can create more problems than they solve if their report opens old wounds, makes strong recommendations on extremely sensitive matters that are not subject to solution at the time, or discusses the wrong problems and issues.

O'Connell and Meeth (1978, p. 41) discuss the following advantages and disadvantages for internal and external evaluators: The external evaluator has the advantages of being competent in program evaluation techniques, having no vested interest in the progrm, and removing the evaluation burden from the existing staff. Disadvantages are that he or she may take longer to understand the program and the evaluation requirements, lacks working relationship with program staff and insitutional personnel, and may be regarded with suspicion by program staff. Advantages of the internal evaluator are that he or she is familiar with the program and staff, understands channels of communication within the institution and its larger community, is familiar with program details, nd may be able to integrate the evaluation into the life of the program. Disadvantages of the internal evaluator are that he or she may not have the skills required for the evaluation, may have a vested interest in the program, and may be overburdened with other duties.

Based upon a report by Pottinger (1975), I devised the following questions to assist in choosing and using consultants:

1. How do you decide that a reviewer/consultant is needed?
   a. What is the nature of help needed?
   b. Is the circumstance specific enough to articulate clearly what is needed in the way of outside assistance? Some possible motiviations include sorting out ambiguous problems, instilling motivation, gaining fresh perspectives, and recommending decisions that internal personnel prefer not to make.
2. How do you choose the proper consultant?
   a. Does the consultant have expertise in your specific area of need?
   b. Several questions might be asked of others who have used the consultant:
      (1) Did the consultant help you further articulate your circumstance without prematurely anticipating a solution?
      (2) Did the consultant help identify resources and approaches within your situation that assisted after departure?
      (3) Did you feel more capable of dealing with the problems after the consultant left?
      (4) Knowing what you know now, would you hire the same consultant if you had it to do over again?
3. How do you most effectively use consultation?
   a. Make one person clearly responsible and accountable for the consultant's work.
   b. Take ample time to define the problem clearly.

c. Explore expectations again about what can and cannot be done. Have a written record.
d. Use the consultant's time effectively. This requires carefully developed visitation schedules.
e. Have a clear understanding of what evaluations and reports are expected or required from the consultant and within what time frame.

3. Self-Studies for Accreditation. Self-study for regional accreditation has been in existence throughout the twentieth century, and it takes place through the six regional accrediting associtions. The Council on Postsecondary Accreditation (1976a, p. 3) gives these historical and current goals of accreditation:

> - foster excellence in postsecondary education through the development of criteria and guidelines for assessing educational effectiveness
> - encourage improvement through continuous self-study and planning
> - assure the educational community, the general public, and other agencies or organizations that an institution or program has both clearly defined and appropriate objectives, maintains conditions under which their achievement can reasonably be expected, appears in fact to be accomplishing them substantially, and can be expected to continue to do so
> - provide counsel and assistance to established and developing institutions and program
> - encourage the diversity of American postsecondary education, and allow institutions to achieve their particular objectives and goals
> - endeavor to protect institutions against encroachments that might jeopardize their educational effectiveness or academic freedom

4. Self-Studies for Institutional Purposes. Hundreds of institutional self-studies unrelated to accreditation have been initiated in recent years, and dozens are underway at any one time. The national trend is toward institutional evaluation with greater rigor and frequency. Institutional self-study can be an important means for self-improvement. The regeneration of the University of Kentucky in the mid-1960's, sparked by a dynamic and experienced new president, came from an extensive self-study, which then became the basis for a comprehensive master plan. Mayhew and Ford (1971, p. 125) mention that "Stanford University shifted its character from that of a strong regional university appealing to bright, wealthy, underachieving students to a university of international stature primarily as a result of the findings of a self-study. Stephens College undertook self-study when its administration believed that the time had come to minimize the traditions of an earlier era. That self-study was used to loosen the soil of academia so that a new president could have a reasonable chance of exercising academic leadership."

The result of self-studies, however, have not been evaluated to any noticeable extent. An exception is the study undertaken by Ladd (1970), which included self-studies done at the University of California, Berkeley; University of New Hampshire; University of Toronto; Swarthmore College; Wesleyan university; Michigan State University; Duke University; Brown University; Stanford University; and two partial cases: Columbia College and the University of California, Los Angeles. The author reached several conclusions based upon firsthand study of these eleven case studies as well as upon analysis of other self-studies:

> Unhappily, the results of these studies seem to lend support--at least in a negative way--to the efficacy of pressure politics as a way of bringing about change. There is little indication in any of the experiences to support the idea that the study-and-report technique is an effective way of gaining acceptance of the need for change or creating enthusiasm for involvement in developing new policies. Where the study-and-report processes were intended primarily to challenge the status quo, they largely failed to do so. When the essential objective was to develop the details of a change in the status quo after the community had already accepted the need for some change, the study-and-report processes were much more effective . . . or where pressures for change from outside the faculties were much in evidence pp. 197-198, 200 .
>
> The cases indicate in both positive and negative ways that strong, skillful leadership is virtually manadatory for the success of any serious effort at educational reform . . . A primary task of academic leadership is to try to counter the pressures favoring the status quo by creating or maintaining an atmosphere of receptivity to change pp. 205-206 .
>
> On the whole the educational policy changes proposed vary considerably in their venturesomeness, and they often seem to speak indirectly--if at all--to the deep malaise that presently affects so much of American higher education. Nevertheless, if the proposals were to be adopted by the institutions concerned, truly consequential changes would be made in the educational policies of those institution. Generally speaking, the cases demonstrate that the proposals developed in the studies became less venturesome or simply diappeared as they passed through the various centers of decision making except where some form of countervailing power was present p. 9 .

The self-studies chosen by Ladd were undertaken between 1966 and 1969, or during the student unrest days. One suspects some studies were undertaken to alleviate the tense and politicized campus climate at the time. Dressel's (1976, p. 409) words may be relevant: "When the only goal of self-study is the alleviation of pressure, the preservation of accreditation, or the attainment of foundation grant, success or failure in attaining the goals often ends the self-study." One can surmise that results of self-studies conducted under these circumstances would be minimal, yet Ladd's conclusions bear careful study.

Research studies have compared studies made of outside, monitored internal, and strictly internal evaluation reports. According to Chambers (1979, p. 32), "these finding show that internal evaluation is as reliable and accurate a method of depicting the current situation as the other two. This does much to dispel the criticism tht the self-study--on its face--is self-serving, biased, and unreliable."

Self-assessment procedures for institutions of higher education also have been developed outside of academe. The American College Testing (ACT) Program (1970) initiated the Institutional Self-Study Service (ISS) to assist a college or university to evaluate the effectiveness of its programs and services. The ACT program was designed to "enable an institution to see itself through the eyes of its students; aid in the quantitative appraisal of college student development; and enable the institution to observe and explore longitudinal trends in student development and opinions on campus" (p. 1). Aspects of student services covered in the survey are faculty advising, counseling, financial needs, extracurricular advising, orientation, housing selections, housing advising, health, and remedial. The ISS survey asks students to indicate the degree of importance they attach to academic vocations and social and nonconventional goods. Instruction, school policies, and physical facilities also receive student evaluation.

The ISS places the student as primary evaluator; and few will disagree with having a significant student input in an institutional self-study, provided students answer questions within their scope of knowledge and experience. The quality of the faculty, academic and institutional leadership, and fiscal management are some essential components in any comprehensive self-study students cannot evaluate effectively. Their opinions may be useful, but data on several key variables in institutional self-study need to be gathered from other constituencies.

Manning (1976), working with the Research and Development Center for Teacher Education at the University of Texas, has developed a troubleshooting checklist (TSC) for higher educational settings to assist those concerned with change in their assessment of organizational variables predictive of an institution's potential for successfully adoping innovations. The TSC consists of one hundred Likert-type items grouped in five categories: organizational change, organizational staff, communications, innovative experience, and student characteristics. The diagnostic and predictive instrument is designed to aid users in estimating the effects of particular variables on the adoption/diffusion process. That is, the TSC provides users with a means of systematically organizing descriptive information in a predictive way. The validity of the TSC remains to be determined, although data on reliability are positive.

The following ten general areas and forty-five evaluative statements form the framework for my system of institutional evaluation (Miller, 1979):

Goals and objectives

1. The goals statement serves as an effective guide for the present and future.
2. Objectives reinforce goals.
3. The institution has adequate planning capabilities.
4. Institutionl admissions policies and procedures are consistent with institutional goals and objectives.
5. The institution's goals and objectives help it maintain a reasonable identity within a statewide system of institutions.

Student learning

6. Students give a good rating to their advising and counseling system.
7. Retention rates are reasonable.
8. An arrary of individualized and compensatory learning resources is available.
9. The student affairs administration is effective.
10. Satisfactory progress toward learning goals is evident.

Faculty performance

11. Current policies and procedures for evaluating individual faculty members are satisfactory.
12. Current instructional improvement/faculty development programs serve their purpose.
13. Faculty personnel policies and procedures are considered satisfactory.
14. Faculty salary scales and fringe benefits are competitive.
15. The overall quality of the faculty's performance is optimal.

Academic programs

16. The institution has effective policies and procedures for developing new programs.
17. The institution has effective policies and procedures for the review and evaluation of existing programs.
18. The general education component is an intellectually stimulating and integral part of the curriculum.
19. The quality and size of the graduate program are consistent with institutional goals and objectives.
20. The library or learning resources center provides good service to the academic community.

Institutional support services

21. The physical plant and facilities are adequate for the size of the student body and for the nature of the academic program.
22. The institution has a relevant and current long-range plan for developing and maintaining its physical plant.
23. Salaries and other benefits for support personnel are sufficient to attract and retain competent individuals.
24. Systematic proceures are used for evaluating the performance of support personnel.

Administrative leadership

25. The administration gives adequate attention to planning.
26. The chief campus administrator and his or her team have effective working relationships with other campus administrators.
27. Institutional governance policies and procedures allow for effective institutional management.
28. The policies and procedures established for administrative evaluation and for professional development are satisfactory.
29. The institution has an effective affirmative action program.

Financial management

30. The tuition and fee structure is compatible with the institution's needs and with the student's capacity to pay.

31. The institution has an efficient management system for accounting and financial reporting.
32. Costs and expenditures are comparable with benchmark institutions.
33. The investment portfolio is well managed.
34. The institution has an effective system for demonstrating its accountability.

Governing board

35. The policies and procedures for conducting board affairs are satisfactory.
36. Trustees understand the differences between policy formulation and policy implementation and apply this knowledge.
37. The governing board works effectively with external constituencies.
38. The board contributes positively to improving the institution.

External relations

39. The institution's activities contribute to the quality of life in its primary service area.
40. The institution has effective relationships with the state higher education (coordinating or governing) office.
41. The institution has an effective relationship with the federal government.
42. The institution is able to secure acceptable levels of funding from private sources and foundations.

Institutional self-improvement

43. The institution seeks improvement through innovation and experimentation.
44. Campus groups have positive attitudes toward self-improvement.
45. The institution has established procedures for evaluting its own effectiveness.

5. State and federal reviews. State educational agencies in the U.S.A. have moved aggressively into the area of program evaluation since about 1972, largely in an effort to stem the tide of new graduate programs that were being developed without careful consideration of their monetary requirements or of long-range institutional goals. New York, Illinois, Louisiana, Tennessee, and others were prominent in their earlier efforts to develop statewide program evaluation efforts.

Tennessee is one state with a sustained effort, with its Performance Funding Project initiated in 1976 by the Tennessee Higher Education Commission. A discussion of the "bold experiement" was reported recently by Ted Marchese. This new element in the state funding formula is that of having five percent incentive funding over and above the regular formula set aside for those public universities that can demonstrate that they are making significant progress toward 14 numeric goals by 1990. These 14 goals are:

- ACT/SAT scores of entrants
- percent of entrants attaining degrees
- "standardized examination scores" of graduates
- scores on the Nationl Teachers Examination
- pass rates on "all parts of professional licensing examinations on the first attempt"

- GRE scores of graduate-school entrants
- the "measured knowledge" of graduate/professional school finishers.

And, as applicable by mission, increases also in:

- gift/grant-based rankings for research activities
- external support for public service programs
- job-placement rated by vocational field
- the match between vocational offerings and service-area job needs
- library holdings.

Plus there is planned to be

- An end to degree-credit remedial courses in public universities, and
- A planned "reduction" in such courses in other institutions (1985, p. 44).

Some recent research has been conducted by Robert Barak, Deputy Executive Secretary of the Iowa Board of Regents, and myself on the role of state level boards of higher education in the review of undergraduate academic program reviews. (I have previously mentioned that the role of state level boards of higher education in review of graduate academic programs is well established.)

Being the first such national study of the undergraduate level, we did not know quite what to expect. We did find that 29 states had some form of statewide undergraduate academic program review and evaluation. This recent trend in the U.S.A. reflects the continuing interest in accountability, the current emphasis on academic quality, and the continuing fiscal constraints on our colleges and universities.

The institutional perspective, however, is different. Just in view of the number of courses and the number of changes the undergraduate program area is much more complicated than the graduate level. And the undergraduate level has been traditionally where the battles over academic freedom have been fought, which means that academics are more sensitive to external influences at the undergraduate level. Perhaps our most disappointing finding was that the state level offices in many states were making little or no use of the results of undergraduate academic reviews to modify and/or improve their programs.

The federal government occupies a complex role in higher education. By virtue of the Tneth Amendment to the United States Bill of Rights, education is left to the discretion of each state. However, many fields such as agriculture, science, medicine, and health have been and continue to be significantly influenced by the federal government. In these areas as well as in a few others, the federal government has been involved in varying degrees of review and evaluation. But generally speaking the major responsibility for the direction and nature of primary, secondary, and postsecondary education is lef to the individual states.

The federal government has been active in a number of other ways tht can indirectly or otherwise influence the quality of our colleges and universities. Examples are federal funds and various grants and loan programs and civil rights and affirmative action policies and mandates. It is not surprising, therefore, that the U.S. government should have an interest in educational quality, as illustrated by the highly publicized book on A Nation at Risk. Federal concern, however, is different from federal control or excessive influence. Institutional and state-level officials are willing to work cooperatively with the federal government in achieving mutually desirable goals of better quality control through cooperation and coordination.

THE ROLE OF PROFESSIONAL JUDGMENT IN THE USE OF ASSESSMEMT DATA

E. Thomas Moran

A significant issue confronting faculty and staff of public colleges and universities is to find effective strategies to utilize Value Added assessment data for the ultimate purpose of optimizing impact on student learning. To accomplish this, assessment information should be effectively fed back into the educational system of a college and used to: 1) strengthen academic advising 2) improve the curriculum 3) enliven teaching, and 4) engage students in more active learning.

To the extent that colleges and universities fail to accomplish these tasks and demonstrate their applicability to enhancing educational quality, the Value Added assessment movement will be preempted by state level policy makers. Individuals in these roles will employ it as a methodology for allocating resources and assuring accountability. The potential dangers of such a consequence are that higher education will become homogenized and its diversity will be diminished, and that highly imperfect measures will be misapplied and result in distorted interpretations of student outcomes and institutional quality.

Avoiding such dangers requires that data be used within institutions to improve the teaching/learning process. This cannot be accomplished successfully unless the professional judgment of faculty and staff is made central to the process of interpreting and utilizing the results of assessments. In this regard. professional judgment is the sine quo non of an effective value-added program.

Two aspects of professional judgment are necessary for utilization of assessment data to occur--a framework for interpreting data and a setting for dialogue. A framework is an analytical design. It provides a context, trends, and comparisons for the data and makes it more relevant and accessible. Wthout a framework, it is often impossible for groups of faculty and administrators to assign meaning and generate actions on the basis of what appears to be an overwhelming mass of unusable information. A setting for dialogue enables faculty to come together and establish relevant, mutually agreed upon interpretations of data, and then to initiate action strategies based on those interpretations. It is this latter issue, the creation of settings and the rationale for doing so, which is the concern of this paper.

THE RATIONALE FOR FACULTY PARTICIPATION

As selected colleges around the country have begun to develop Value Added assessment programs, their attention has been directed largely toward the problems of measurement. Accordingly, questions related to the selection of instruments have dominated discussion on campuses. These discussions apparently proceed under an implicit assumption that utilization of the right measurements will produce data that, by their very existence, will result in institutional improvement. But, experts have observed that even when data are available, they are rarely relied upon to improve institutional functioning. (See Lawler et al, 1985, for a discussion of this issue). Furthermore, an emphasis on measurement overlooks the central role of professional judgment on the part of faculty and staff in interpreting data and in forging and implementing action plans based on those data. (For a comprehensive discussion of professional judgment in the creation of knowledge, see Lindblom and Cohen, 1979.)

The problem confronting the faculty and staff of Colleges is to build strategies to maximize the utilization of Value Added assessment data for the ultimate purpose of optimizing impact on *student learning* and promoting institutional change. As noted, to accomplish this, assessment information should be effectively fed back into the educational system of a college and used to 1) engage students in more active learning 2) enliven teaching 3) improve the curriculum where necessary and 4) strengthen academic advising. Seen in this context, both the *results* of assessment as well as the *process* of utilizing assessment data are critical to a Value Added approach.

It is important to stress that Value Added is not simply an effort to evaluate student and institutional performance, but a method of achieving student involvement, establishing realistic expectations, and providing timely and relevant feedback on student performance, particularly learning and development over time. The N.I.E. study group report, entitled *Involvment in Learning*, established the conceptual foundations for this orientation. It emphasized the importance of student involvement, appropriate expectations for student learning, and timely feedback on performance as crucial elements in attaining educational excellence. (Benjamin Bloom's Book on *Human Learning* offers much empirical support for these concepts.)

Value Added is also a method of assisting faculty to use teaching strategies and classroom activities that encourage more active learning on the part of students and which emphasizes assessment as part of the learning process (See Zelda Gamson's book entitled *Liberating Learning*. Viewed in this regard, the assessment and evaluation components of a Value Added project are only a first step in achieving a larger and infinitely more important purpose--that of engaging students in ways that will result in their achieving greater gains in learning and development. In an analogous manner, it is also a method of involving faculty more fully in the central purposes of the college.

A primary concern, however, is that for these assessment strategies to be transformed into activities that actually promote student development, rather than merely measure performance, appropriate methods for the utilization of assessment data must be developed and employed by faculty and staff. These individuals must collectively find ways to use assessment in strengthening teaching and advising, as well as curriculum development, and in clarifying institutional identity. There are no existent, pre-programed methods that can simply be borrowed by colleges for these purposes. Value Added is far too new a concept for a fully developed paradigm to have emerged which can be imported by any college.

Moreover, an important aspect of assessment is that it serves as a stimulus to engage faculty and administrators in a process of institutional scrutiny. (See Peter Ewell's work entitled The Self-Regarding Institution for a oomplete discussion.) Indeed, the data from Value Added assessments are merely suggestive rather than conclusive. It is only through the collective interpretation of assessment information that evaluation and initiatives for improvement can occur. Therefore, the utilization of Value Added assessment for the purposes identified above, must be developed locally.

Two elements are necessary for utilization of assessment data to occur--a framework for interpreting data and a setting for dialogue. A framework is an analytical design. It provides a context, trends and comparisons for the data and makes it more relevant and accessible. Without a framework, it is often impossible for groups of faculty and administrators to assign meaning and generate actions on the basis of what appears to be an overwhelming mass of unusable information. A setting for dialogue enables faculty to come together and establish relevant, mutually agreed upon interpretations of data, and then to initiate action strategies based on those interpretations.

Up to this point the discussion has focused on uses of assessment data that are internal to an institution. However, there is also a larger background against which these issues should be understood. For reasons that will be explained below, public colleges and universities are more immediately effected by developments in the area assessment because legislative action employing student testing to induce change can be so readily directed at them. (See The Education Commission of the States, "Levers for Change" by Peter Ewell.)

In the past eighteen months, some state level policy makers have begun to preempt student assessment data and consider ways to design and use it to "rationalize" resource allocation among institutions of higher education in their respective states. (See, for example, the Tennessee Model. More recently The Northeast Missouri Value Added project has spawned a similar approach in that state). Rational allocation schemes, of the type currently being proposed in these states, employ pre-specified criteria to evaluate and compare institutions on the basis of student performance data and to allocate resources based on formulas that can be derived from such evaluations.

While this general approach emerges out of a basic intent to improve accountability in higher education, it has several dangers. Firstly, it politicizes data on student learning and development, and thereby increases the likelihood that the use of student assessments on campuses will be more guarded and less innovative.

Secondly, the actual measures (assessment instruments, etc.) under consideration are inevitably imperfect as are all measures of this type. They have technical properties that dictate their use for carefully limited purposes--and none of these instruments has been designed for inter-institutional comparisons.

Thirdly, any widespread use of common assessment instruments can result in homogenization of curricula on campuses which would likely diminish the rich diversity of American higher education.

Fourthly, assessment of student learning and development is a process that is highly dependent on the professional judgment of educators. Student growth is a complex phenomenon that can only be fully understood when considered along with such factors as student background characteristics and specific institutional purposes.

Fifthly, comparisons of institutions based on student assessments may produce a new, but, highly problematic hierarchy of institutions. Standardized tests are normative and, therefore, force variation in scores among respondents. This means that some individuals must inevitably be ranked above the norm and an equal number below the norm. (Criterion, reference tests give everyone taking the exam a theoretical chance of performing equally because they report scores on the basis of right answers rather than basing standards on comparisons with others who have taken the test.) The problem, then, with normative instruments, which are the typed generally being considered by state-level analysts, is that, when applied to populations of institutions, they invite a hierarchy of institutions based on gross measures of student performance. This runs the risk of substituting one troublesome set of hierarchical standards, those based on inputs (e.g. student SAT's, or high school G.P.A.'s) for another which is based on student outcomes. In its pure form, a hierarchy based on student outcomes is to be desired, but, because of technical difficulties, it is likely to be created on the basis of formulas that do not adequately consider differences in students incoming characteristics or aspirations, or differences in missions of specific institutions. (Nevertheless, if the unhappy alternative is to continue to accept the current hierarchy of institutions based only on resources, then on based only on student outcomes, however imperfectly measured, has the advantage of being based on values that are consistent with the democratic and egalitarian purposes of American public higher education.) (See Alexander Astin's recent book, Achieving Educational Excellence) for a discussion of models of quality in higher education).

Despite these problems, state level policy makers may exhibit increasing tendencies to use student assessment data to insure accountability and allocate resources, unless institutions begin to collect these data themselves and demonstrate that they ar appropriately utilizing them to improve the impact their institutions have on students. Such efforts on the part of colleges will constitute implicit measures of accountability, while minimizing the potentially misleading results that may occur by pooling student assessment data at statewide or other inappropriate levels of analysis.

An additional reason for making the professional judgments of faculty and staff central to the utilization of Value Added assessment data is that research demonstrates that faculty involvement is critical to the success of curriculum reform efforts. Indeed, faculty participation has deep historic roots; in most institutions faculty either control the curriculum and its attendant learning processes or exercise the greatest influence upon them (see Faculty Participation in Decision Making by Carol Floyd).

In summary, the task facing institutions of public higher education is to forge a model that can be adapted by institutions in their own efforts to incorporate assessment data into settings where these data can be coupled with professional judgment and employed to focus on educational improvements. To the extent that such a model can be developed, it will obviate the need for these data to be employed at state levels where they are subject to greater misapplication. Even more importantly, devising effective uses of assessment data can help prompt change efforts at organizational levels within institutions where those efforts can have their most immediate and beneficial impact on students.

Bibliography

Astin, Alexander. *Achieving Educational Excellence*. San Francisico: Gossey-Bass, 1985.

Bloom, Benjamin S. *Human Characteristics and School Learning*. New York: McGraw Hill, 1976.

Ewell, Peter. *The Self-Regarding Institution: Information for Excellence*. Boulder, CO: National Center for Educational Management Systems, 1984.

Ewell, Peter. *Levers for Change*. Denver, CO: Education Commission of the States, 1985.

Floyd, Carol E. *Faculty Participation in Decision Making: Necessity or Luxury*. ASHE-ERIC Report number 8. Washington, D.C., Association for the Study of Higher Education, 1985.

Gamson, Zelda. *Liberating Education*. San Francisco: Gossey-Bass, 1984.

Lawler, Edward, E. et al. *Doing Research that is Useful for Theory and Practice*. San Francisco: Gossey-Bass, 1985.

Lindblom, Charles K. and Cohen, David D. *Usable Knowledge*. New Haven: Yale University Press, 1979.

Mortimer, Kenneth et al. *Involvement in Learning*. Final Report of the Study Group on the Conditions of Excellence in American Higher Education. Washington, D.C.: National Institute of Education, 1984.

E. Thomas Moran is Assistant Vice President for Academic Affairs at SUNY Plattsburgh, Keho 313

THE PROMISE OF ASSESSMENT: LINKING INDICATORS TO OUTCOMES

Thomas E. R. Redmon

It is difficult to be pessimistic about the ultimate outcomes of education at any level in the United States. The recommendations and rhetoric of some two dozen national reports on plans to improve the opportunities to learn and the commitments to teach in our colleges and schools suggest that all is well. And yet, the disarray of the current moment is evident enough to persuade most parents, students and teachers not to view education as their sole source of personal and career development. Indeed, if the nation's investment in education were situated in stocks and bonds, there could be no doubt that the bears would be in control and that the general climate of education would represent the near bottom in a depression that began not too many years ago.

The Promise of Assessment

Assessment appears to be not only a useful term in our education dictionary but also our only way out of an otherwise unmeasurable decline in the benefits of this most treasured American entitlement. Recently, the term has gained credibility, highlighted by a 1985 meeting in Columiba, South Carolina of educators reponsible for assessment, sponsored by the American Council on Education. Whatever gain may be realized from this nation's climb out of a depressive educational era appears to be linked closely with the effective use of assessment as a tool to measure our future progress.

Traditionally, assessment has been used to make reference to any number of measurable scales of progress. The measurable progress might be related to achievement or to some other structural improvement or change in institutions or in state responsibility for education.

Three examples come to mind as the most widely used assessment activities of current interest to institutional decision makers. First, the increasing use of tests of academic performance is reaching an all-time high on college campuses. These tests of academic ability are used at the admissions level all the way through "exit" levels from the sophomore or senior year of college. Perhaps the tradition of classroom evaluation lies at the heart of this increasing use of performance measurement. Certainly the concept of testing students on their achievement and ability does not require extensive explanation or defense in faculty senates or other campus decision groups. Our abiding faith in objective (and now essay) testing contributes to this long-standing, traditional method of assessment.

The use of such tests requires a number of important decisions about the choice of an institutionally developed instrument or the application (and perhaps modification) of a generally available commercial product. Regardless of the choice, conscientious administrators and faculty have learned what it takes for such testing to produce useful outcomes. A brief review of such experience indicates a number of requirements for any assessment by way of objective or essay testing of academic performance:

1. Clear purposes regarding the overall use of tests for particular application as well as the choice of a particular instrument to measure defined achievement or ability.

2. Clear question specifications allowing the matching of measures to expected outcomes.

3. Comprehensiveness of the instrument to adequately measure ability or achievement in the areas of concern.

4. Explicit connections to classroom settings, instructional method and course structure.

5. An informed decision regarding the choice of essay and/or objective items as the forms of evaluation.

6. Clear directions students will understand without lengthy explanation.

7. A well-articulated policy relating the outcome of test scores to overall evaluation by grades.

8. Opportunities to evaluate the test according to the difficulty of items on the instrument as well as the validity of the instrument for the purpose defined.

Although many of these requirements are taken into consideration at most campuses, assessment is not always consistent and the use of such criteria becomes burdensome in many settings of evaluation.

Additional concern also should be focused on at least three areas of validity for such instruments:

* Construct validity (making sure the idea of using a test fits the circumstances for evaluating academic outcomes)

* Content validity (making sure that the evaluation instrument chosen or created fits the specific instructional and curriculum setting within which students are to be evaluated)

* Predictive validity (making sure that the test will serve the purpose of certifying preparation for the next step of education in a specific setting)

A second typical area of assessment that has evolved to judge the effectiveness of college may be found in many assessments of student satisfaction with college. Although the evaluation is administered typically in the form of a survey (with an infinite variety of such instruments used at hundreds of colleges across the nation) the primary purpose of such an effort is to ascertain the level of satisfaction students have with their undergraduate experience. The quality of such assessment generally depends upon the sophistication of the survey research design and the expected use of the outcomes of the survey. Noel and Beales (1979) found significant factors relating to the retention of students in a follow-up study of student satisfaction with their undergraduate experience. Their findings suggest the value of such assessment and its impact on institutions with significant retention problems. Specifically, their study yielded five major negative factors which students find to be unsatisfactory.

1. Inadequate academic advising
2. Inadequate curricular offerings
3. Conflict between class schedule and job
4. Inadequate financial aid
5. Inadequate counselling support systems

Conversely, the study also found certain factors to be positively related to students' satisfaction with their undergraduate experience.

1. Caring attitude of faculty and staff
2. High quality of teaching
3. Adequate financial aid
4. Student involvement in campus life
5. High quality of advising

Regardless of whether institutions use these two examples of assessment or other traditional forms of measuring student satisfaction and ability, there appears to be great comfort taken in simply doing the testing or administering the survey. The real value (and difficulty) of such assessment is in the ability of a college or state system of higher education to effectively incorporate assessment into the best possible educational outcomes for students.

There is, of course, a third "inferential" use of assessment. State systems of postsecondary education increasingly aggregate and summarize information on student satisfaction to obtain a larger view of the outcomes of public education policies. But the effects of assessment often have little to do with the desired outcomes embodied in state plans and institutional missions. As a comprehensive example of the kind of outcomes that assessment should yield, the Southern Regional Education Board's recommendations through its Commission on Educational Quality are worth careful review.

In 1985, the SREB Commission published 50 specific goals in three important areas of educational reform:

1. Providing access to quality undergraduate education
2. Improving teacher education in colleges and schools
3. Setting academic standards for secondary vocational education

These three areas of educational reform not only organize useful groups of educational outcomes but appear (on the basis of recent responses) to provide direction for educational activities in the most critical (depressed) arenas of educational quality.

Since the Truman Commission on Higher Education and the Harvard "redbook," countless institutional, state, and federal reports have called for moderate to dramatic tinkering with what had become known as undergraduate general education. The current spate of reports offered to various audiences to confirm that rigorous self study is underway continues that tradition of appraisal from within and without. The SREB goals call for specific actions, or, if you will, outcomes, in key problem reas.

> Setting standards for college-level academic performance. (Including specific recommendations on entry-level standards and indicators of satisfactory progress throughout the undergraduate course of study)
>
> Making college remedial programs accessible to students. (Including recommendations regarding both the geographical and financial accessibility of college preparatory study)
>
> Preparing students for college. (Including strongly worded recommendations for increased cooperation between high schools and colleges, and the development of explicit measurements of adequate preparation for college-level work)
>
> Strengthening the college curriculum. (Specifying goals for content reform in undergraduate curriculum and underlining the importance of faculty cooperation in reform activities)
>
> Preparing faculty for teaching undergraudate studies. (Highlighting the importance of faculty attitudes and skills related to teaching undergraduates during the first two years of college work)

The commission Report on Improving Teacher Education included a large number of goals related to four major agendas for higher education and the schools:

> Recruitment of talented students. (Including concerns for adequate academic preparation of students and appropriate incentives to attract students into education for the teaching profession)
>
> The content and structure of professional education. (Twelve of the goals in this report focused on the program structure, the content of study and the practical experience that form the substance of professional education for teachers)

Institutional support. (Including goals on making teacher education improvement a priority agenda item for presidents and chancellors, increasing instituional support for faculty in colleges of education, and advocating increased contact between professional teacher educators and public school administration and teaching)

State support. (Including concern for public support for teaching, state executive support and increased legislative funding for teacher education programs)

Finally, the Commission's Report on Improving Secondary Vocational Education included goals in two major areas:

Maintaining quality and raising academic standards. (The goals included concern for restructuring the linkage between vocational and academic education, course design, basic skills assessment and remediation, and the important relationship between practical experience and academic learning)

Experimenting with new models. (The recommendations suggested the formating of state task forces to encourage pilot projects to improve secondary vocational education, increasing attention to evaluation methods in vocational education, joint enrollment opportunities, upgrading the use of technology in vocational education, and exploring new structures of cooperative education)

These goals in three key areas of education provide a foundation on which effective assessment may be constructed. Each of the categories discussed above demand assessment tied closely to useful results rather than general knowledge about progress, if the outcomes of education are to benefit students.

While many authors assume that the outcomes of education are related directly to measurable scales on tests or other evaluation instruments, this paper will suggest the larger definition of an educational outcome as a goal related to the overall purpose of assessment in any specific setting. For state and institutional policymakers, the definition of an educational outcome as a measurable statemtn of a goal yet to be accomplished provides two important foundations for research and policymaking:

1. The outcome of a particular educational activity such as assessment must be perceived as a clear statement of what ought to be (that is, an outcome must give a sense of direction for the future).
2. The definition of outcomes as specific, measurable goals should allow for any number of evaluation or testing techniques to be used to determine if progress is underway or if the outcome has been achieved.

Such a definition allows outcomes to be used more directly to formulate educational policy and to make use of more than one specific indicator of progress or attainment. Further, this view allows educational outcomes to be studied with respect to their costs and benefits to individuals and institutions and to the society that supports them. As Howard Bowen (1977) has cautioned, educators can make a significant mistake "by assuming that improved outcomes are desirable regardless of cost ... In assessing the efficiency of higher education, ... one must consider both cost and result." (p. 21)

Only in this critical perspective will assessment of academic ability, student satisfaction, and state level progress goals give policymakers the information necessary for improvement.

The results of assessment have never been easy to interpret and have been even more difficult to put into action plans. While it is unrealistic to assume that decision makers in higher education will be using both the skills of psychometric interpretation and administrative policymaking, there is a way to connect the results of assessment to outcomes.

Analyzing the Costs and Benefits of Assessment

The administrative tool of policy analysis is now well-defined enough for good application to this difficult problem of linking assessment results to intended outcomes. Five basic steps of policy analysis may be useful to administrators seeking to make better use of assessment. Although there are many models of policy analysis in use, the following adaptation of five steps suggested by Stonkey and Zeckhauser (from A Primer for Policy Analysis, 1978) provides the basic idea for using policy analysis.

1. Defining the problem and its context. Much of the ultimate failure of assessment programs may be blamed on the small amount of time given at the beginning to determine why assessment is needed in the first place. Some critics of assessment would even suggest that with a little thought and consideration at the beginning, many assessment programs would never be started. But assessment, as other technical tools, tends to be easily adopted as a proven method of documenting conditions and the need for change.

2. Developing alternative courses of action. Before specific assessment instruments are used to study academic progress, student satisfaction, or state progress toward educational goals, administrators would do well to outline various alternatives for getting the work done. Tests and assessment may be only one of many alternatives and, indeed, may prove to be the best. But "the best" must be defined carefully--not only in terms of time, staff, and money--but also in terms of anticipated results and uses. Many institutions in states make premature assumptions

about alternatives, assuming that their choice lies in picking one of the many models of assessment in use or in theory around the country. There may be, however, other approaches to solving the problem that do not involve testing in the form of commercially available instruments designed for general use but not well formulated for specific problems in institutions or states. Even if ultimately one instrument is chosen to help solve the problem, the critique involved in considering alternatives will make administrators far more aware of the limitations of test instruments and will help policymakers to judge the limited utility of those tests within a larger context of evaluation and policy decision.

3. Predicting the consequences of alternative courses of action. A few states and many institutions are now feeling the effect of assessment policies. Predicting these effects is an important part of choosing an adequate assessment policy. Strategic planning, forecasting, and environmental scanning represent a few of the sophisticated approaches to this step, and are no longer feasible except in corporate centers with large computers. Most analysis of consequences in postsecondary education, however, is still done on the basis of opinion and campus interest group pressure.

4. Evaluating outcomes. Before choosing a course of action, administrators must evaluate the various effects of alternative courses of action. This is a difficult step. Most institutions are satisfied to have listed the consequences of various alternatives and rely upon a committee on individual intuitive judgement regarding the value (positive or negative) of a course of action. This step of policy analysis requires a more formal, explicit statement of the relative value of those outcomes. In the long run, this step will give those who must later consider changing a plicy a clear sense of the intentions underlying the final policy choice at this earlier stage.

5. Choosing a course of action. This step is inevitable, but is oftentimes taken without great confidence. The how and why of the final decision is important for its impact on individuals and groups that have devoted time and energy to making the decision clear and on those who have been involved in analyzing the potential consequences of the decision. If faculty and staff have been involved in any of the steps described above, there will be a strong interest in hearing the results of their work. At times, the final decision will be quite clear and all that remains is good timing for its promulgation. At other times, the decision will be painful and will involve holding some interests (and the individuals who have those interests) at

a higher priority than others. The main point of this step in policy analysis is to insure that the decision about assessment is not arrived at by default or accident.

Despite the documented utility of policy analysis for considering decisions related to assessment in higher education, some final comments are appropriate on one aspect of choosing assessments.

Planning and Controlling Outcomes

As in any organizational system, there is no reason to spend scarce resources on outcomes that do not help fulfill the goals of the organization. Underneath the interest in higher education institutions for calm, neutral research, colleges and states want something good to happen to the students who enroll. Much of that good can be found in the statements of goals with which institutions will become committed in various ways over the next decade. This paper will conclude with a call for concern not so much for the inevitable choices of desired outcomes but for increasing commitment to plan for and control those outcomes in order that they will, in fact, occur.

In the 1985 ACE meeting on assessment in Columbia, South Carolina, Professor John Harris of David Lipscomb College presented an excellent overview of assessment. He introduced his presentation through a letter to Dr. Dubious Scholarius, Dean of the Faculty and Vice President for Academic Affairs at Everyone University, Oxbridge, U.S.A. In his contrived letter, he recognized Dr. Scholarius' skepticism about "the whole outcomes business," but went on to assure the Dean of how important his interest in this area can be to the mission of Everyone University.

In John Harris' closing paragraphs, we find reassurance for the Dean at Everyone and a bit of advice:

> As you move into assessment at Everyone, I suggest you find someone to observe and comment on the organizational development implications of what you want to do. My point is this: An emphasis on assessment will affect the way everyone functions as an organization. You not only need help in the technical side of assessment but also in the nurturing of a climate characterized by:
>
> 1. Deep concern for "results" over "form."
> 2. Commitment to high standards.
> 3. Concomitant interest in helping students reach the standards.
>
> In the last analysis, an emphasis on assessment is more of an attitude than a collection of tests. (p. 53)

Harris' advice to the Dean should be well taken. All of the technical improvement possible in test making and administration will not necessarily result in the outcomes of education that we all might expect or hope to have happen. Taking the time to plan for the outcomes we expect will add enormously to the success of whatever evaluation or assessment colleges undertake to move closer toward their overall goals.

Although planning models are widely available in written as well as software form, such effort often is wasted becuase of the lack of control over the actions and efforts to bring about desired outcomes. Institutions often times create elaborate, sophisticated plans on paper without effective management controls to bring about results.

As institutions and states undertake plans for assessment that will yield useful outcomes, certain requirements of control are necessary:

1. A rational, organized approach to planning and to the choice of desired outcomes is essential to any accomplishment expected.

2. Clear documented decisions on what to control will assist planning and will aid in the accomplishment of the goals and objectives institutions and states set for themselves. These decisions should especially represent consensus on what is viewed as critical in the accomplishment of the outcome and what forecasts trouble in the movement toward that goal.

3. The senior management team at an institution or in a state must decide on the method of control that will be used to guide and encourage progress toward the desired outcomes. Such decisions are especially important in higher education where individuals are used to some degree of anonymity and creative participation rather than routine line responsibility.

4. The mangement team must also decide when to control efforts toward setting goals and choosing assessments. These decisions can effectively stifle or encourage the accomplishment of outcomes at both the institutional and state level. In education the decision to control is in many ways a matter of timing.

5. Progress toward outcomes must itself be measured in order that all participants may see clearly that their efforts are leading toward desired outcomes.

In view of the increasing sophistication and availability of high quality assessment instruments, the importance of linking tests to outcomes should not be underestimated. Outcomes such as those in the areas described above are not attained easily by even the most coherent software analysis package. Time and effort afforded to planning and control cannot help but improve the attitude of "assessors" and "assessees." Even more practically, however, such an emphasis may ensure that assessment is worth its cost.

Thomas E. R. Redmon, Senior Officer, Educational Policy Research, Southern Regional Education Board (SREB), 1340 Spring Street, Atlanta, Georgia.

EVALUATION FOR EXCELLENCE: THE TENNESSEE PLAN

Linda Rudolph
James Nixon

Tennessee has received recognition across the nation during the past few years for its emphasis on excellence in education. New ideas such as the career ladder plan for elementary and secondary teachers and centers of excellence and chairs of excellence in higher education have resulted in controversy, acclaim, recognition, and improved quality at all levels. The Comprehensive Education Act of 1984 has stimulated exciting changes for all levels of education in Tennessee.

Of particular interest to institutions of higher education were the five-year legislative goals developed by legislators. Following are the benchmarks to be obtained within a five-year period (1989) by all institutions of higher education in Tennessee (quoted directly from Section 97 of the Comprehensive Education Act of 1984).

(1) An increase in the percentage of students who enter four-year university degree programs and who subsequently earn baccalaureate degrees.

(2) An improvement in the average National Teachers Examination scores of students enrolled in public university teacher preparation Programs.

(3) An increase in the scores of public university entry-level students on the composite tests of ACT and SAT.

(4) An improvement in standardized examination scores of graduating seniors at public universities.

(5) An increase in the number of students from public universities who pass all parts of professional licensing examinations on the first attempt in fields for which a licensure examination is required.

(6) An improvement in test scores of students entering graduate schools within public universities as measured by such national examinations as the GRE.

(7) An increase in the measured knowledge of graduates of public university graduate and professional programs.

(8) An improvement in the library holdings of the public technical institutes, community colleges, and universities.

(9) For those universities whose defined role includes research an improvement in the ranking of the public universities' research activities as measured by additional external grants and gifts received for sponsored research (recognized, however, that changes in federal research policies are beyond the control of individual institutions or the State of Tennessee).

(10) An improvement in the support given to public universities' public service programs as measured by additional external funds received for such activities.

(11) An improvement in the job placement rate by specific vocational fields studied for all vocational graduates of area vocational schools, technical institutes, and community colleges.

(12) An improvement in the correlation of specific vocational fields of study offered by area vocational schools, technical institutes, and community colleges with the specific vocational needs of each service area of the state as determined by projections of the State Departments of Planning, Employment Security, Economic and Community Development, and Labor.

(13) The implementation by public universities of policies which ensure that no credit is offered for courses which provide remediation for high school deficiencies will apply toward minimum degree requirements for graduation.

(14) A reduction in those courses now offered for degree credit by public technical institutes and community colleges which serve as remediation for high school deficiencies (recognizing that until such time as the basic skills possessed by all high school graduates in Tennessee are significantly improved, these institutions will continue to have a significant remediation responsibility).

Progress toward the accomplishment of these benchmarks is reported to the governing board of the institution each year, which in turn compiles the data for an oversight committee of legislators. This information has been used by legislators for budget appropriations to institutions. Specific needs have been highlighted; specifically in the case of Austin Peay State University (APSU) the institution has received special monies for recruiting minority population and for programs to assist under-prepared students. Other recommendations such as an increase in admission standards to decrease attrition have been implemented upon recommendations resulting from this data. It is expected that data gathered concerning these bench marks will be used increasingly by legislators in decision making relative to the needs of higher education.

In addition to the legislative goals described above, the governor and legislature demanded improved evidence of the quality of teaching and of the curriculum from higher education. No longer is the public satisfied with the unsupported statements about the quality of education and the advancements being made in the area. This demand for improved quality and evidence thereof has resulted in a massive evaluation movement in the state with built-in rewards through performance funding for gains and improvements in various areas of instruction.

After completion of an intensive research project, the Tennessee Higher Education Commission in 1983 defined five areas of instructional evaluation. Colleges and universities meeting the point criteria assigned to these five areas can earn up to 5 percent of the instructional component of the institution's formula-generated education and general budget. The areas included attainment of accreditation for eligible programs, evaluation of academic offerings and general education outcomes, evidence of improved instruction, and an institutional plan for instructional improvement. Following is a discussion of each variable as described by material published by the Tennessee Higher Education Commission.

Variable I--Program Accreditation

Programs meeting or exceeding the standards of acceptable accrediting bodies may be awarded up to 25 points. Points are awarded based on the percentage of eligible programs accredited. Acceptable accrediting bodies include COPA members or those approved by the governing boards as acceptable. At Austin Peay State University, we earn points for the accreditation of 18 out of 23 eligible programs. The only programs not presently accredited are Chemistry, Art, Social Work, Industrial Technology, and Business. Progress is being made for accreditation of all programs at the direction of the President. New programs, under five years old, are not required to meet the conditions of this variable.

This is the only variable for which Austin Peay has not obtained full funding the past four years. There is some concern in Tennessee universities as to whether the return for accreditation under the performance funding formula balances the cost for the accreditation process. However, many believe that accreditation brings positive returns to programs through other means, such as increased attractiveness of the program to students. Seeking accreditation for all programs presents other problems since many programs are quite small and the standards of accrediting bodies quite rigorous, and since the content focus of some accrediting agencies differs from the content focus of specific university programs.

Variable II--Program Field Evaluation

Variable II, program field evaluation, rewards institutions that evaluate the quality of each academic program at least once within a five year cycle. Additional points may be earned under this variable for institutions which demonstrate increasing quality or above average quality based on test results. The quality of academic programs may be measured by externally validated tests, locally developed tests, or external peer review. A plan for the external peer review which includes the vita of two evaluators, a schedule of the review process, and a provision for a written report must be submitted to the Tennessee Higher Education Commission for approval prior to the review. Locally developed tests must be constructed with the assistance of two external consultants or cooperatively with other similar institutions. Again, the Tennessee Higher Education Commission must approve the construction plan for a locally developed test prior to the activity. Externally validated tests must also be approved by the Tennessee Higher Education Commission Prior to use. All three evaluative methods have been used successfully by APSU academic programs. Chemistry, physics, math, history, and biology have relied on the Graduate Record Examination; however, this test is a costly method for evaluating large numbers of senior students. In addition, some programs have found that the emphasis of the GRE test in their area does not match the emphasis of their program. For example, the History Department at APSU focuses on American History whereas the GRE test emphasizes World History. Psychology, agriculture, public management,sociology, social work, geology, and philosophy have worked jointly with sister institutions to develop local tests. This procedure is being described by another presentor. Programs in art and drama have invited external peer reviewers to evaluate their offerings.

For assessing a representative sample of graduates, an institution may earn up to 10 points under this variable. Institutions may earn an additional 20 points if they demonstrate that the performance of their graduates exceeds the norm of an externally validated test, or the performance of their graduates has improved over that of the last tested group on a validated or locally developed instrument. Peer review does not generate the additional 20 points under Variable II. APSU has earned maximum points under this variable for the last four years.

In addition to the points earned toward performance funding, departments assessing their academic programs have benefited from the program review in other ways. The Psychology Department reports that a review of the scores of their graduates by areas indicated strengths and weaknesses of the major. A revision of the major is currently in process and changes in the program based on test results will be reflected. The Math Department and a number of the sciences have used the results effectively for revision of their curriculum. One departmental major was terminated on the basis of test scores.

Variable III--General Education Outcomes

Four-year institutions and community colleges have been directed by the Tennessee Higher Education Commission to use the ACT-COMP test (objective or composite forms) for assessment of general education outcomes. ACT reports that their goal was "to design, develop, validate, and implement assessment instruments and procedures to measure and evaluate certain kinds of knowledge and skills that undergraduates are expected to acquire as a result of general education--the knowledge and skills that are believed necessary for successful functioning in adult society" (ACT, 1980).

Five points may be earned for the assessment of graduates. An additional 20 points can be earned if the valued-added performance of the graduates is above average when compared to the scores of comparable institutions,or if the institution shows gain over its previously reported value-added measures. Value-added is calculated by comparing entering ACT scores with existing COMP scores, estimating the entering COMP scores based on a statistical table, or value-added may be measured through longitudinal or cross-sectional test data.

The ACT-COMP measures general education outcomes in six basic areas: . Functioning in Social Institutions, Using Science and Technology, Using the Arts, Communicating, Solving Proglems, and Clarifying Values. The results have yielded interesting and informative data over the past six years (discussed in presentation). Additionally, the areas measured are similar to the recommended required areas of emphasis for general education as suggested by the Association of American Colleges (February, 1985): (1) Inquiry, abstract logical thinking, critical analysis; (2) Literacy: writing, reading, speaking, listening; (3) Understanding numerical data; (4) Historical consciousness; (5) Science; (6) Values; (7) Art; (8) International and multicultural experiences; (9) Study in depth. Since Austin Peay is currently in the process of revising its general education core and the entire curriculum in preparation for a move to the semester systam, the data from the test provide valuable suggestions for curriculum development. The reports provided by ACT interpret the results for the institution and suggest additional areas of study.

Variable IV--Instructional Improvement Based on Referrent Group Survey

In order to provide for evaluation of program quality by the consumer, the Tennessee Higher Education Commission developed a method for rewarding institutions which survey students, alumni, employers, or community members. For simply conducting the survey and reporting the results plus planned improvement actions (if necessary), 5 points may be earned by the institution. An additional 5 points may be earned if the institution conducts two or more surveys of referrent groups, or administers the survey more than once and shows improvement. The survey must yield quantifiable data, be nationally or locally constructed, and must be approved by the Commission prior to administration.

APSU has chosen to meet this criterion through the use of two survey instruments: student evaluation of instruction and ACT alumni survey. From the student evaluation of instruction, a report analyzing the results outlining indicated action from these analyses, and describing actions taken is submitted to the governing board. The ACT alumni survey is administered systemwide and the outcomes compared to group norms. It has been interesting to note that the results from the ACT alumni survey have confirmed the data yielded from the COMP test.

Variable V--Planning for Instructional Program Improvement

In order to link evaluation to the planning for improvement within the institution, a variable to reward institutions for these efforts is included in the performance funding program. An institution may earn up to 5 points for submitting an annual plan which focuses on improvement in instruction and includes specific goals and benchmarks or measurable objectives which reflect the mission of the institution. Faculty involvement in this process is mandated. An additional 5 points may be earned for the submission of evidence indicating that at least half of the objectives and benchmarks for the previous year were met and favorably evaluated.

As a result of the planning process specific goals for instruction at APSU have been developed. They include supporting the work of an academic advisement task force for three years charged with the mission of thoroughly studying the present advisement system and developing a plan for improved advisement in order to decrease attrition and improve the academic environment. Additionally, on campus forums for the presentation of professional papers and review of recent writings in a variety of fields were planned to stimulate interdisciplinary interest. Methods for recognizing and encouraging outstanding instructional performance were developed. Efforts to train faculty in test construction and understanding assessment measures were organized. These are only a part of the numerous ideas that have resulted from cooperative planning for instructional improvement.

As previously stated, an institution may earn up to 5 percent of its institution's formula-generated education and general budget. APSU is a small regional university with a student population of approximately 5,000. During the past year APSU earned .0465 of an approximately $15 million education and general budget. An increase of $697,500 can add significantly to a small instructional budget.

Conclusion

Legislative involvement in assessing and directing higher education causes fear and trepidation for many of us in higher education. The Tennessee plan has brought with it many advantages and some pitfalls. Overall, I believe that most educators in the state believe that the returns have been most profitable, not only financially but also in the improvement of education in the state.

References

Association of American Colleges (February, 1985). Integrity in the college curriculum: A report to the academic community. Washington, D.C.

Comprehensive Education Act of 1984, Section 97, 5-year legislative goals. Tennessee Higher Education Commission.

The American College Testing Program (1980). COMP Prospectus for the institutional use of COMP materials in assessing general education knowledge and skills.

Dr. Linda Rudolph is Acting Dean of the College of Graduate and Professional Programs, Austin Peay State University, Clarksville, Tennessee 37044.

Dr. James Nixon is Dean of the College of Arts and Sciences at Austin Peay State University, Clarksville, Tennessee 37044.

TEACHING THINKING:
IMPLICATIONS FOR INSTITUTIONAL AND PUBLIC POLICY

Mark A. Schlesinger

Preface

This paper is about irony. It suggests that faculties, institutions, and state policy makers have not necessarily thought well about teaching college students to think well, and recommends how the above agencies might think better on the matter. I would prefer to focus on guidelines for policy rather than criticize specific examples of such policy. The reader is free, of course, to draw his or her conclusions about extant testing programs and state curricular goal sets.

The path winds this way: the paper explicates briefly and broadly one model of teaching thinking, and examines what faculty would have to do to employ the model effectively; the model and its role prescription for faculty form the basis for a discussion of the posture and actions of institutional and supra-institutional policy makers. In sum, the paper asks that faculty and policy makers invoke the sound habits of problem solving we wish our students to adopt.

Departure

We start with a premise that should find no disagreement among administrators, legislators, coordinating board members, and faculty members: fundamentally, a college education should assure that the graduate think well. Without that premise, we are lost. As with many things, it is often most convenient not to define explicitly what we mean by "think well." The concept falls in the category of Quine's (1964) "promissory notes," terms upon which we agree not to agree, for to force agreement would be to breed dissolution.

However, when an entity is flailing around and causing itself great harm, it is time to encourage exploration into, if not total agreement on, the meaning of its basic premises. The exploration will not culminate in a uniform sense of what we mean by "think well." As happens when Achilles must try to catch Zeno's Tortoise, "the game. . . will have to be played an infinite number of times . . ." (Hofstadter, 1980, p. 32). However, the essence of scholarly inquiry is that we play the game together, cooperatively and cumulatively. In the pages to follow, I shall suggest institutional and supra-institutional approaches to change that can encourage such inquiry, and can approximate a model of intellectual sophistication and problem solving acumen we would hold up to our students -- those whom we hope to teach to think well.

The Significance of Teaching Thinking

Clearly the move to more explicitly teach thinking skills to college students carries deep significance for students and, we are told, society (Paul, 1985). But I think of a different kind of significance, one less frequently and grandly verged in the literature: the meaning such a move has, or should have, in the workaday life of the faculty member. Interestingly enough, while presumably we have always taught students to think, the literal commitment to monitor whether the students in fact learn to think well, whether the skills they learn matter now and later, whether they become more disposed to use their intellectual skills -- this commitment to do what we have always said we do and now recognize we don't do is as radical as is the resolve in any other context to ask, "what are we supposed to be doing, anyway?" The question, in this instance, seems to have many answers (Schlesinger, 1984). Rather than cover them here, I take a particular approach to teaching thinking skills that my department has employed. It combines a perspective on intellectual development with theory on problem solving.

Intellectual Development. Paradigms charting human cognitive development abound. Such models differ in important ways. However, most provide a view of intellectual activity -- not necessarily capacity -- that progresses from simple to complex, from tunnel vision to contingent vision. As an example, the Perry (1970, 1981) "scheme" has deservedly attracted much attention. It portrays student growth from a dualistic view of the world (black and white, right and wrong) through the apprehension of multiplicity and the struggle to manage relativism to the ability to make commitments while mindful of their contingency. Perry does not argue that all individuals progress through these stages; he shows clearly how humans may shy away from commitment, crippled by relativism. At the same time, he shows how humans may, at less sophisticated levels of development, leap to commitment untempered by relativism. I ask that you keep such progressions in mind as we examine the proper approach of policy makers to facilitate teaching students to think well.

Too, one should note the following: it is now apparent that no scheme of cognitive development makes sense when viewed as a uniform life progression. I hope I am able to function at Perry's upper levels when considering scholarly issues in my field. But confront me with a leaky faucet and I'm dualism (perhaps Perry should have called it "duelism") personified. We must all go through "developmental" stages every time we learn something new. In this sense, skill learning is "recursive." And this recursiveness becomes even more apparent when we view learning as problem solving.

Problem Solving. The growing attention of psychologists to internal thought processes has resulted in a wealth of literature on problem solving. Recent work suggests that problem solving may be an appropriate term for thinking in general (Anderson, 1980, p.259). Such a view gives a different slant on other operations we call "thinking." Failure to perform higher order operations (e.g., the kinds of tasks we commonly portray as "analysis" or "synthesis" or "critical thinking") may not be so much a failure of intellect but the outcome of "a lack of knowledge about appropriate problem solving strategies and techniques" (Anderson, p. 237).

My department begins by teaching explicitly the following skills that underlie effective problem solving (Woditsch, 1977):

- The conscious, selective directing of one's attention;
- The suspension of decision or judgment; reflectiveness;
- The separating of large, complex tasks or arrays of information into manageable bits, and the meaningful recombining of those bits;
- The forming of analogs, considering the new in light of what is known;
- The continual checking of one's progress, especially when a solution seems to offer itself; "strong inference" (Platt, 1964) -- a disposition to falsify as well as to confirm.

Adults can -- and at some low functional level must -- learn such skills (Whimbey, 1975). They may be tacit in sophisticated problem solving, in which the individual familiar with the situation, possessing the appropriate information, and committed to the endeavor evinces such facility. But place that same individual in an unfamiliar situation, with informational and task cues unclear, and the skills of which I speak must be summoned from the background.

A Synthesized View of Development and Problem Solving. The truly competent problem solver is one who is disposed to invoke such skills at the upper reaches of what we have commonly called intellectual development. The right side of Figure 1 represents, after a fashion, such a context. In fact, our courses and our activities within courses progress from the left side of the chart to the right. And our thesis continually proves itself: to deal successfully with the right side of the chart -- those problems which the solver must define, with unspecified tasks and ambiguous or nonexistent information -- require that the individual invoke the skills I have mentioned.

Finally, we apply the following principles concerning the manner in which we most effectively assist adults to become effective problem solves:

- The student learns best by engagement in solving problems, and reflection on such engagement. Doing it is more productive than being told, a priori, how to do it (Schlesinger, 1984, and many others before).
- Problem solving thrives as a social activity. Engagement with and criticism from peers are powerful aids to problem solving (Lochhead, 1979).

These principles are empirically as well as commonly sensible. They are behind the awakening sentiment that faculty . . .

> . . . should make greater use of active modes of teaching and require that students take greater responsibility for their learning. (Mortimer, et al., p. 40)

EASIER ¼-- MORE DIFFICULT

Amount and Location of Information

need little information	need much
already possess much	possess little
much supplied	little supplied
must find little	must find much

Nature of Information

internally consistent	inconsistent
consistent with student's experience	anomalous
clear	vague, ambiguous
stable, predictable	volatile

Nature of Task

(perceived as) relevant	irrelevant
familiar	unfamiliar
clear	vague, ambiguous
simple	complex
unitary	multiple

Figure 1 A scheme for viewing and sequencing problems

The Faculty Imperative. This model of effective problem solving, if it is valid, applies beyond the classroom. For the moment, let us employ it as a template, and shift it from the classroom to the problem solving context involving the faculty member who must reconfigure instruction explicitly to teach students to think well. Consider what is required of this individual. He or she must represent the discipline not as a collection of facts, theories, and concepts, but as a problem solving domain. Its uniqueness lies in the manner in which its professionals specifically apply their broad problem solving skills. Perhaps some among the remarkable gathering of higher education reformers in Wisconsin last fall had such a disciplinary role in mind when recommending that the faculty "give more attention to assessing the major, with academic departments asking themselves, 'What are we trying to achieve?'" Jacobson, 1985). Such questions cannot be answered, let alone meaningfully asked, in an atmosphere that presses for immediate and concrete programs. But more on that pressure below.

As the Wisconsin conferees also recommended, our faculty member will have to "inventory the skills that undergraduates need"; more, he or she must locate what he or she can contribute to those skills. But here is where I begin to get nervous, as I hope you do, too.

How to create such an inventory? After all, we are dealing with the workings of the intellectual mind, something even psychologists avoided until recently. And we seek to bridge a tenuous link between our understanding of those workings and the classroom, a context in which those workings might be made whole and effective. Truly, this is an area in which faculty know very little. Its problems are not defined, let alone its solutions. Its information is a mix of aphorism, discredited lore, contradictions, and finally a trickling of findings which promise more. In short, the problem context is about as vague as it can be.

And yet we have many such inventories. It is well to participate in their derivation, for such participation renews one's commitment to the enterprise and marks our progress. But the state testing programs, the goal rosters, and even institutional competence lists as refined as Alverno College's run a great risk that I shall call the "outcome fixation."

The Outcome Fixation as Problem Solving Model

I have discussed a model of thinking well, incorporating notions of intellectual development and problem solving, as it guides instructional practice at one college. I have begun to extend the model, using it to depict faculty who seek substantive reform (e.g., to use this particular instructional model); they are as students in a problematic context. It may stretch matters somewhat to add institutional and state level policy makers to the extended model in a role analogous to teachers, but policy makers resemble faculty to the extent both wish to encourage those who receive their services to take ownership of a problem, to reflect patiently, and to work cooperatively and cumulatively. In short, they wish those who must solve problems to think well. Within this brand of thinking well, one finds the roster of problem solving skills noted above, and one witnesses those skills at a high level of intellectual functioning. It may be most helpful to present general examples of the converse approach.

Solve it all at once. Policy makers not thinking well in an unpredictable and ill defined context (e.g., reforming higher education to teach thinking skills more effectively) would pose and expect solutions for an entire problem rather than its constituents. They would expect rapid change, and would reward change that was on its face most comprehensive. Thus, the individual faculty member experimenting with a better way of, let us say, helping physics students understand proportionality, would attract little or no attention, while the institution with an entirely new curriculum or goal statement would reap rich reward. In a situation with a clear need for prolonged deliberation (the translation, say, of an entire profession such as teaching into a roster of competencies), institutional groups of faculty might be given a year or two!

Add to this rush to judgement a proclivity to stick with that judgment, and you have yet another indicator of policy that neither reflects nor encourages thinking well. The temptation in the current educational and political climate is to consider goals and tests as more than measures of accomplishment -- they *are* the accomplishment itself. Thinking well would dictate a continual re-examination of not only those goals and tests but the premises on which they are based.

Going for the whole solution has the consequences you might intuitively expect, and which experts on problem solving have also explicated. In a matter as complex as curricular outcomes, focusing on the "whole" solution ignores or obscures implications for pedagogy, for administration and support, for organizational implementation, for assessment, and for revision or even re-examination of founding precepts. How frequently have sponsors of a test or goal set invited efforts to disprove it? The political environment encourages answers and promises, not successive approximations.

Outcome vs. Process. In the classroom, a focus on outcomes may preempt attention to something more important (at least in the domain of problem solving skills): the manner in which the student achieves the outcomes. A paper, for example, is a solution to many problems, including how to write the paper itself. But it is an incomplete indicator of the writer's skill, as the writer may have arrived at his/her solution in any number of ways, including imitation (Woditsch, 1975; Schlesinger, 1984).

Curricular reform frequently focuses similarly on outcomes to the detriment or exclusion of process. An emphasis on goal statements without support for the process by which human beings must negotiate meanings that underlie such statements would be a fine example of poor problem solving. It would be compounded to the extent such goal statements became their own reality, rather than hypotheses or recognized approximations of answers in constant search of better questions.

The irony is most apparent in efforts to teach thinking skills. If ever there was a hunt whose meaning lay in the pursuit rather than the kill, this is it. Even more than organizational goals, curricular goals acquire meaning not in their stated form but in individual and institutional behavior (Woditsch, 1973).

Recommendations

Those who make or influence curricular policy can promote substantive reform in many ways beyond the charting of goals and the writing of tests. They can influence the preparation of college teachers; they can encourage on-line faculty to ponder their work; they can bring together critical groups of such faculty; and they can abet assessment clearly aimed at improving the model. Such recommendations are resoundingly familiar. But their strict implications -- as the strict implications of teaching thinking skills -- may not be so familiar.

Preparation to Teach in College. As the Wisconsin conferees noted, graduate schools can more effectively prepare their students for college teaching. But more important than featuring educational psychology and classroom management, preparation should engage the future professor in the debate over the meaning of his or her discipline, particularly for undergraduates who may take only one course therein. Hence, sponsorship of programs that train future professors should occur in the context of the wider effort to examine instructional mission.

The Faculty Member as Epistemologist. Legislators and state agencies should view program development as just that -- emphasizing the action connotation of "development" -- rather than as program establishment. Too, they should promote clarification of questions rather than immediate provision of answers. In this spirit, they might encourage individual faculty members to examine their assumptions, to form and test new hunches, and, most importantly, to do so in collaboration with other faculty. (For an example of such activities, see Katz, 1985.) In promoting and rewarding little steps, sponsors could provide specific funding for outside research expertise, release time for planning and assessment, working papers that are permitted to be tentative and even wrong, and activities that bring the experimenters together (e.g., consultant visits, special issue conferences). Additionally, states might award funds for faculty to spend extended time -- a semester or a year -- in "active reflection" on their teaching and discipline. The Shaughnessy grants formerly offered by FIPSE were an illustration of such a program.

Such efforts must be entirely in the spirit of assisting the faculty to solve its own problem, and they must be seen in that light. If they are linked to future funding, program evaluation, and especially personnel issues, the impetus for faculty members to reconsider their teaching -- what is, after all, the core of their professional life -- will be deflected by at least two kinds of reactions I once ran into at a small college whose faculty had received a directive to "modularize" the curriculum. One faculty member told me, "Modules? They're great. I've designed 150 of them this year." Another, in confidence, said, "I don't know what the hell a "module" is, and I'm going to sit right here until they tell me."

Pluralism and Unity. Policy makers face conflicting pulls of pluralism and unification. While they encourage faculty to examine their instruction individually and in small groups, they must address a countervailing need to help integrate such efforts. Pluralism by itself is little more than anarchy. The kind of integrative unity of which I write is not one of operational goals, but a unity of commitment to understand better the instructional mission. Rather than imitate much of the formal research on education, which frequently is replicative rather than cumulative, our faculty-experimenter-researchers should be encouraged to learn from each other, and even to contradict each other. Again, toward this end, states can back ventures that bring faculty together around common problems. The CUE Center at Bowling Green State University, an early FIPSE beneficiary, performed this function by assisting faculty with disparate definitions of critical thinking, for example, to test those definitions in the classroom against common criteria.

A Conclusion on Reflections

All of the above recommendations can take place in the context of program development. They do not require a splendid isolation, and thus should not drag down institutional efforts to respond quickly to the shifting forces of the environment. They demand, however, that the response be cause for reflection, that faculty be encouraged to invoke yet another trait of the effective problem solver: a disposition to check, and even to prove one's most desired outcomes false. This is perhaps the hardest aspect of effective problem solving to invoke in a politically charged setting. As good problem solving instruction allows students to be wrong and encourages them to learn from it, so should sound policy allow faculty and institutions to be wrong and learn from it. The alternative is to be wrong and to stay wrong.

Regardless of specific ways in which faculty and institutions receive encouragement to bring about substantive curricular reform, there must also be an attitude favorable to reasonable risk taking and the avowed failure. If in fact this is too quaint a hope, then much of the above is likewise. For reform thrives when the consequences of failure are to try again, rather than to be left out the next time.

References

Anderson, J. R. Cognitive Psychology and Its Implications. San Francisco: W. H. Freeman and Company, 1980.

Hofstadter, D. R. Godel, Escher, Bach: An Eternal Golden Braid. New York: Vintage Books, 1980.

Katz, J. Teaching Based on Knowledge of Students. In J. Katz (Ed.), Teaching as Though Students Mattered. New Directions for Teaching and Learning, #21. San Francisco: Jossey-Bass, 1985.

Jacobson, R. L. Leading Advocates of Reform in Undergraduate Education Find that It's Not So Easy to Move from Rhetoric to Action. Chronicle of Higher Education. October 9, 1986, 24.

Lochhead, J. On Learning to Balance Perceptions by Conceptions: A Dialogue Between Two Science Students. In J. Lochhead and J. Clement (Eds.), Cognitive Process Instruction: Research on Teaching Thinking Skills. Philadelphia: Franklin Institute Press, 1979.

Mortimer, K. P. & Others, Involvement in Learning: Realizing the Potential of American Higher Education. Report of the Study Group on the Conditions of Excellence in American Higher Education. Washington, D. C.: National Institute of Education, 1984. Reprinted in Chronicle of Higher Education, October 24, 1984, 35-49.

Paul, W. R. The Critical Thinking Movement: A Historical Perspective. National Forum, 1985, LXV, 2-3, 32.

Perry, W. G., Jr. Forms of Intellectual and Ethical Development in the College Years: A Scheme. New York: Holt, Rinehart and Winston, 1970.

Perry, W. G., Jr. Cognitive and Ethical Growth: The Making of Meaning. In A. W. Chickering & Associates (Eds.), The Modern American College: Responding to the New Realities of Diverse Students and a Changing Society. San Francisco: Jossey-Bass, 1981.

Platt, J. R. Strong Inference. Science, 1964, 146, 347-353.

Quine, W. V. O. Word and Object. Cambridge, Massachusetts: MIT Press, 1960.

Schlesinger, M. A. The Road to Teaching Thinking: A Rest Stop. Journal of General Education, 1984, 36, 182-196.

Whimbey, A. Intelligence Can Be Taught. New York: Dutton, 1975.

Woditsch, G. A. Paradigmatic Thought: Cognitive Behavior for Governance in Higher Education. Ph.D. Dissertation, University of Michigan, 1973.

Woditsch, G. A. Developing Generic Skills: A Model for Competency-Based General Education. Bowling Green, Ohio: CUE Project Publications, Bowling Green State University, 1977.

Mark A. Schlesinger is with the Department of Analysis & Communication at the College of Management, University of Massachusetts-Boston, Boston, Massachusetts 02125

THE USES OF ASSESSMENT FOR DECISION MAKING: A PRIMER ABOUT NORTHEAST MISSOURI STATE UNIVERSITY'S VALUE-ADDED PROGRAM

Terry B. Smith

Northeast Missouri State University is a state-supported institution that was founded as a teachers college in 1867 and was designated Missouri's public liberal arts institution in 1985. The enrollment of 6,500 is approximately 90% undergraduate and is two-thirds Missouri resident students.

Early in his presidency Charles McClain envisioned a national surge toward quality and accountability in higher education. In 1973 he started, incrementally and at first on a voluntary basis, an assessment program that on several dimensions and at several stages tapped the intellectual value added by classroom instruction at Northeast. The key components are the nationally standardized tests that are given toward the end of the liberal arts portion of the curriculum (the sophomore test) and the exams given during the semester before graduation (the senior test.)

Sophomore Tests

There are two sophomore test models. Virtually all freshmen bring ACT results with them. Half the sophomore class is given an ACT retest to determine individual and cohort learning growth in English, Math, Social Science, and Natural Science. The results of this test/retest show increases --usually-- in achievement on the four major subject areas assessed by the ACT (Figure 1.).

Early results in Math were discouraging, as shown in figure 1; scores actually declined between the freshman and sophomore year. Faculty, stimulated to further inquiry by these data, discovered also that student surveys revealed a low rating of "understanding mathematics" and that senior students were performing poorly on quantitative portions of senior tests.

To address these deficiencies several curricular changes were proposed and adopted. The mathematics requirement for General Education was increased from "math appreciation" to college algebra, business and accounting majors were required to take calculus, and a no-credit math lab was introduced.

Students were advised into a stronger math sequence and close attention was paid to changes in outcomes of tests and surveys. Math scores began to increase both university-wide (Figure 1.) and in the division that had even further upgraded requirements (Figure 2.), and student satisfaction about understanding and applying math rose (Figure 3.).

The other half of the freshman class, having taken the ACT-COMP (College Outcomes Measurement Program) during the first semester of the freshman year, retakes the ACT-COMP during the second semester of the sophomore year to determine individual and cohort learning growth in six liberal learning skill areas.

A steady increase in "value-added" as measured by the ACT-COMP has been reported since this assessment began in 1980 (Figures 4. & 5.). Simultaneously student attitudes about their liberal learning experiences improved (Figure 6.).

Senior Tests

Seniors take the senior test appropriate to their discipline, typically the GRE and NTE, to assess whether students graduating with degrees in those areas are nationally-competitive (defined by NMSU as scoring above the 50th percentile) (Figure 7.). A number of disciplines do not have nationally-normed senior tests; local tests are devised as substitutes.

The Psychology program at NMSU is an apt illustration of the use of senior tests in evaluating curriculum at the upper division level. Psychology was a popular major in the late 1970's and early 1980's, with large numbers of majors and enrollments and high gpa's in the major. Scores on the Psychology GRE, however, showed limited subject mastery; fewer than one senior in four was testing at the 50th percentile.

Faculty ultimately concluded that the major was lacking rigor; seniors with low GRE's had found the "soft track" through the curriculum and were avoiding core courses (developmental, learning, social psychology). Faculty then reduced electives, strengthened requirements, and upgraded "enrichment" courses.

Psychology GRE scores increased appreciably (Figure 8.), and while it would seem that Psychology majors might be less satisfied with a more demanding program, the reverse turned out to be true (Figure 9).

The Individual Student Level

Northeast's assessment model permits informed diagnosis, advisement, and tracking of an individual student, on which there are enormous usable data (Figure 10.). Using only entering ACT data a student can be advised into an early liberal arts curriculum that will correct deficiencies and enrich strengths. One such student (Figure 11.) experienced considerable growth in communication skills because of careful placement during his freshman year. His sophomore ACT retest validated the wisdom of informed course placement (Figure 12.).

Other Data Use

In addition to using test scores and survey results for student advisement and curriculum development, Northeast has used assessment data to justify appropriation requests to the state's coordinating board, governor, and legislature. Because the assessment program is, among other things, our accountability tool, the state has given the University several special appropriations based on needs uncovered by the data; e.g., more faculty in disciplines where value-added goals were not being met (Figures 13. & 14.).

Finally, many non-instructional units have used not just the survey but also the test data for program development and improvement. For example, a number of scholarship awarding and renewal decisions hinge on sophomore test performance and high sophomore test scorers get preference for the most desirable residence hall spaces.

The most sublime use of the Value-Added asessment data is to raise universally the awareness level of the entire community about the major enterprise of the institution: student learning and achievement. Everyone contributes to a student's capacity for intellectual growth, not just the teaching faculty.

Safeguards and Watchwords

Essential to a successful assessment program is two-way trust. Administrators or external agencies must not use data to make punitive programmatic, personnel, and budgetary decisions, but rather must use data to encourage curricular improvement and faculty development. Cross-discipline or isolated-faculty comparisons are patent violations of trust.

At the same time, faculty should acknowledge that these data, while not perfect, are superior to anecdotal information and unreferenced grading practices. When combined with other more traditional inputs assessment data permit meaningful, grounded change in curriculum, teaching style, advisement, etc.

Standardized achievement tests are valid and reliable instruments when carefully and appropriately used. They are most powerful as change agents when combined with other sources of information; the "triangulation" effect permits a faculty and administration to measure real growth with marked precision (Figure 15.).

Student attitudes about assessment are, of course, critical. Colleges can take several steps to increase student support for assessment and consequently their motivation to perform at peak levels:

- avoid add-on "testing fees;"
- tell students about benefits of the assessment program while recruiting them;
- continue to market the benefits of assessment during the academic year to enrolled students;
- reward students (publicly, if authorized) for strong sophomore and senior test performances and for significant "value-added" shown on the sophomore test;

* be sure the test-taking environment is optimal (quiet rooms, good time of day, good lighting, cheerful proctors, etc.);
* provide guarantees of confidentiality of scores; * get periodic feedback from students about the assessment process;
* generally, do not make testing a punitive experience in any way.

After all, the ability to make managerial decisions based on the results of assessment depends on the level of faculty and administrator confidence in the validity of the scores.

Conclusion

Value-Added assessment focuses on individual student achievement and on the commitment of resources necessary to maximize that achievement. It involves the entire institution and clearly isolates student learning as the primary activity of the academy.

It provides a feedback loop for academic and support services planning. It can underpin performance-based external funding requests. Most sublimely, Value-Added assessment provides a high-level mode of accountability and demonstrates for all to see a commitment to guaranteeing the integrity of the baccalaureate degree.

Terry B. Smith is Professor of Political Science and Dean of Students at Northeast Missouri State University.

FIGURE 1.

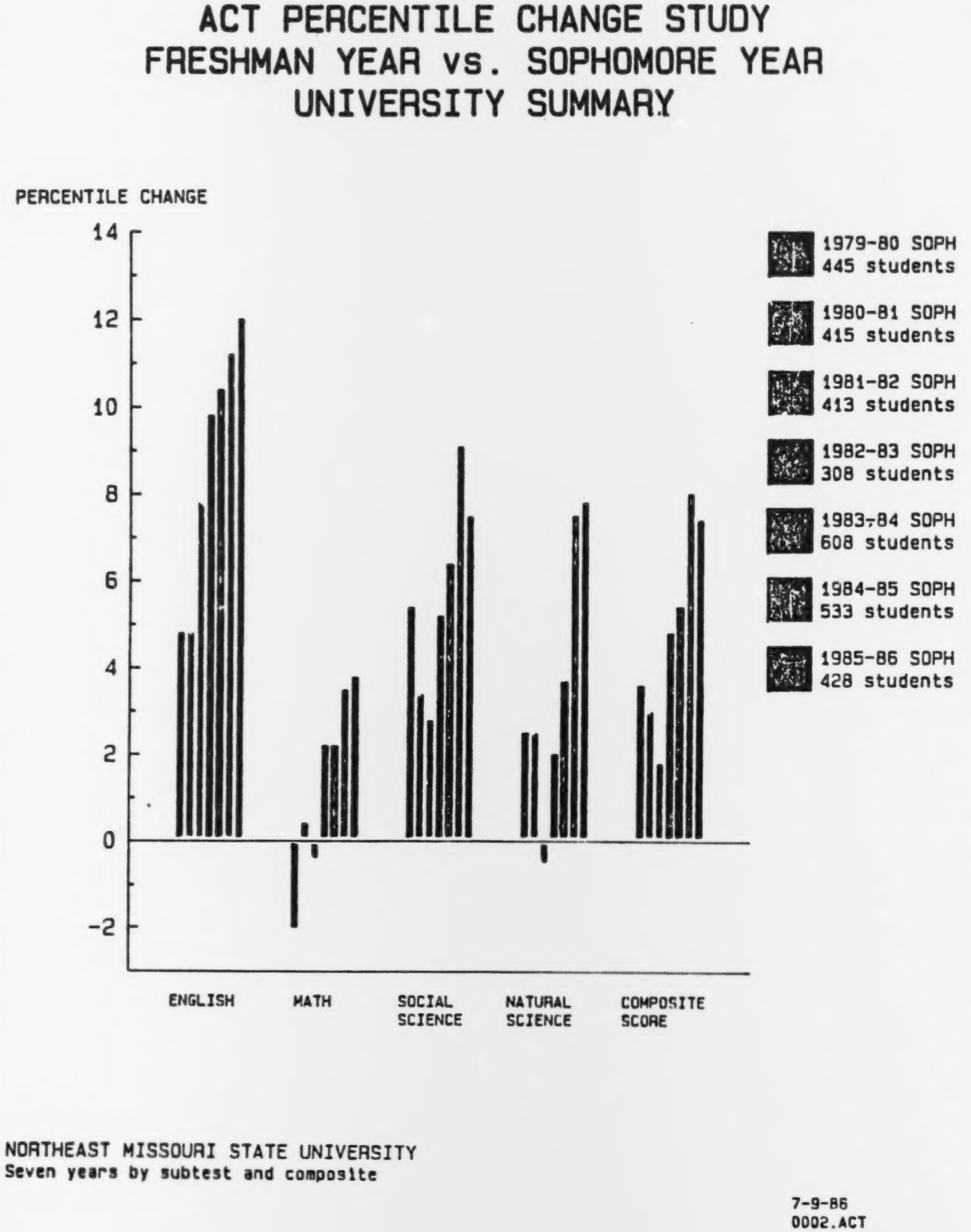
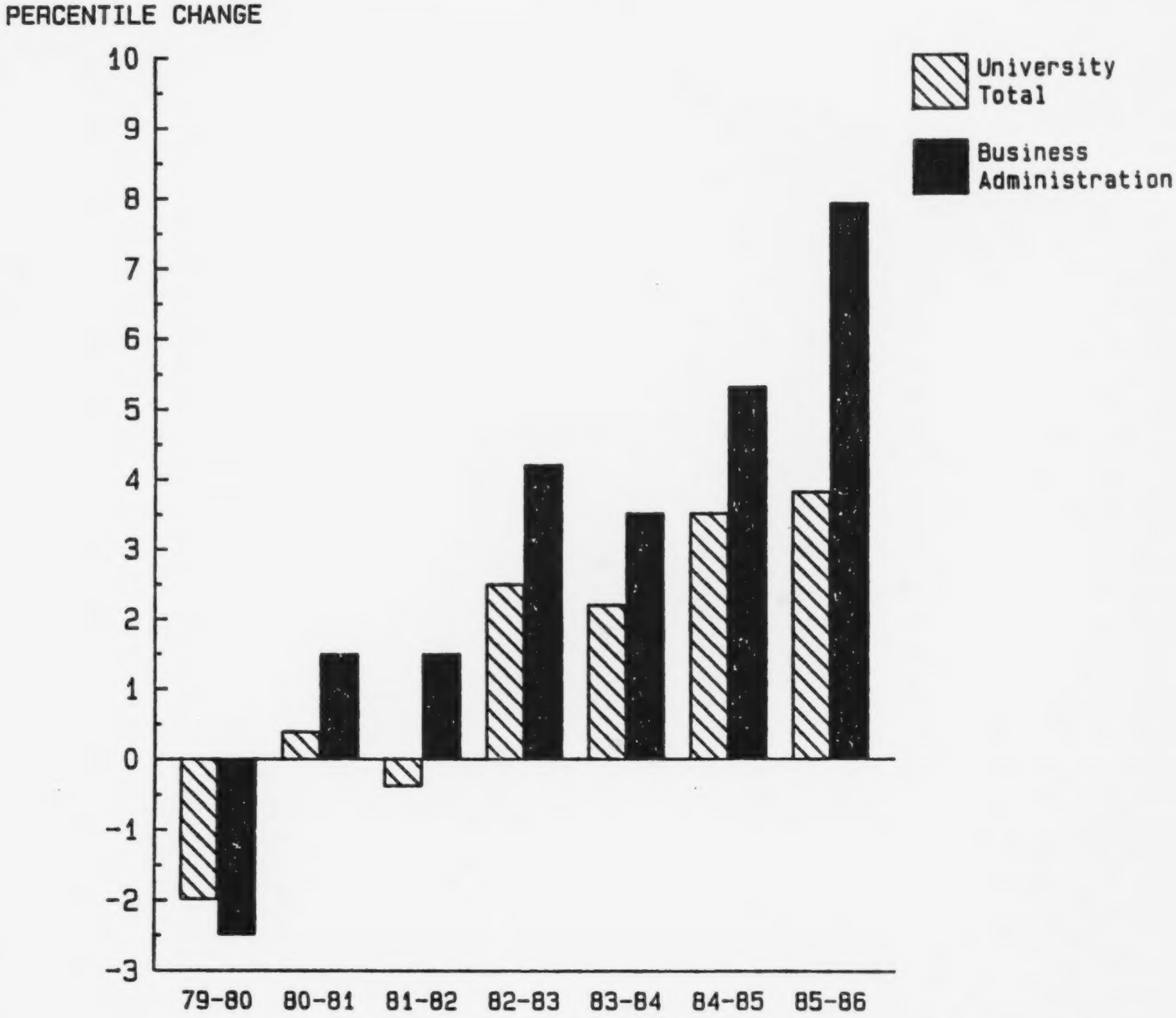

ACT PERCENTILE CHANGE STUDY
FRESHMAN YEAR vs. SOPHOMORE YEAR
UNIVERSITY SUMMARY
PERCENTILE CHANGE
14
12
10
8
6
4
2
0
-2
ENGLISH
MATH
SOCIAL SCIENCE
NATURAL SCIENCE
COMPOSITE SCORE
1979-80 SOPH 445 students
1980-81 SOPH 415 students
1981-82 SOPH 413 students
1982-83 SOPH 308 students
1983-84 SOPH 608 students
1984-85 SOPH 533 students
1985-86 SOPH 428 students
NORTHEAST MISSOURI STATE UNIVERSITY
Seven years by subtest and composite
7-9-86
0002.ACT

FIGURE 2.

MATH ACT PERCENTILE CHANGE STUDY
Freshmen vs. Sophomores
University Total vs. Business Administration



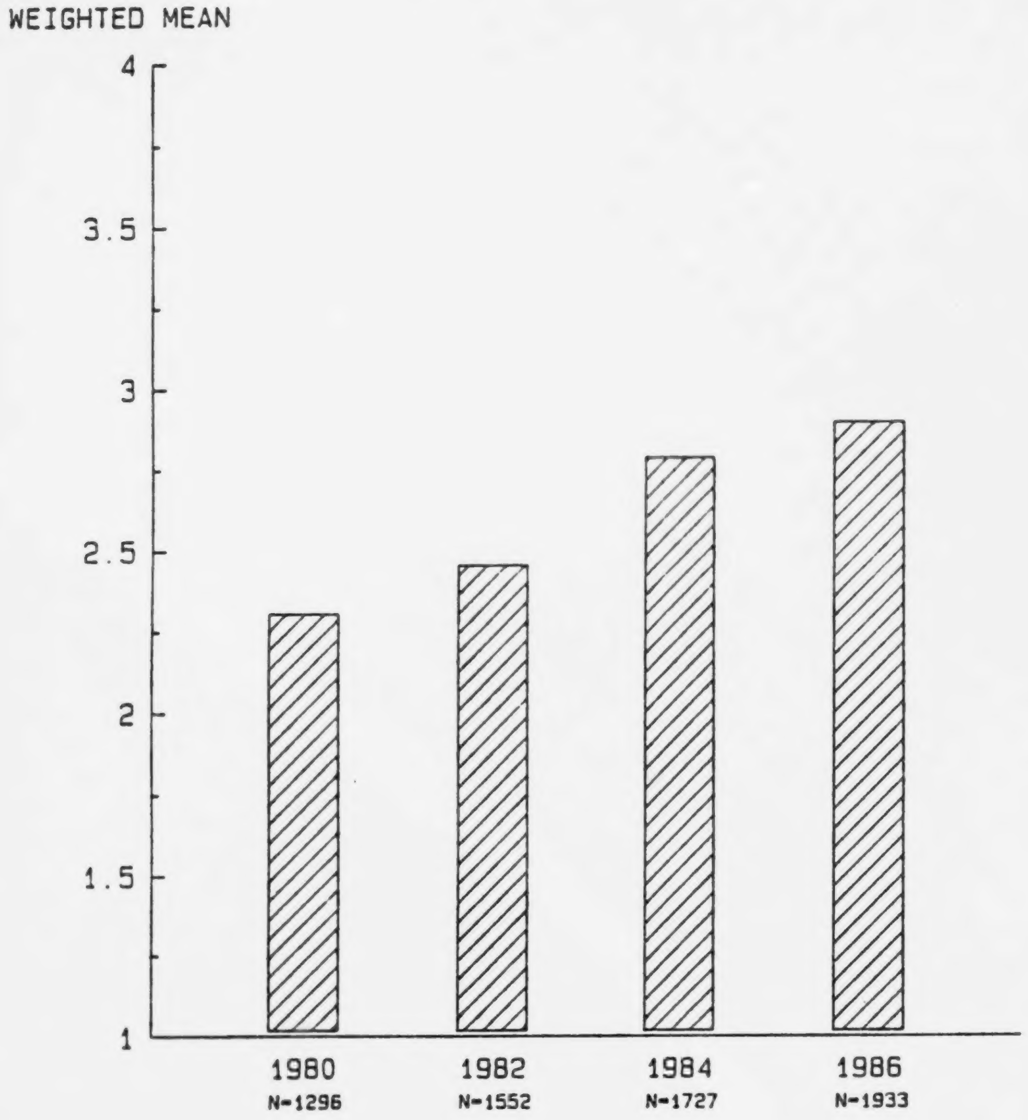
NORTHEAST MISSOURI STATE UNIVERSITY
Five years by math subset

7-14-86
0010.ACT

FIGURE 3.

INSTITUTIONAL STUDENT SURVEY

Weighted Means On Student Satisfaction Understanding And Applying Mathematics

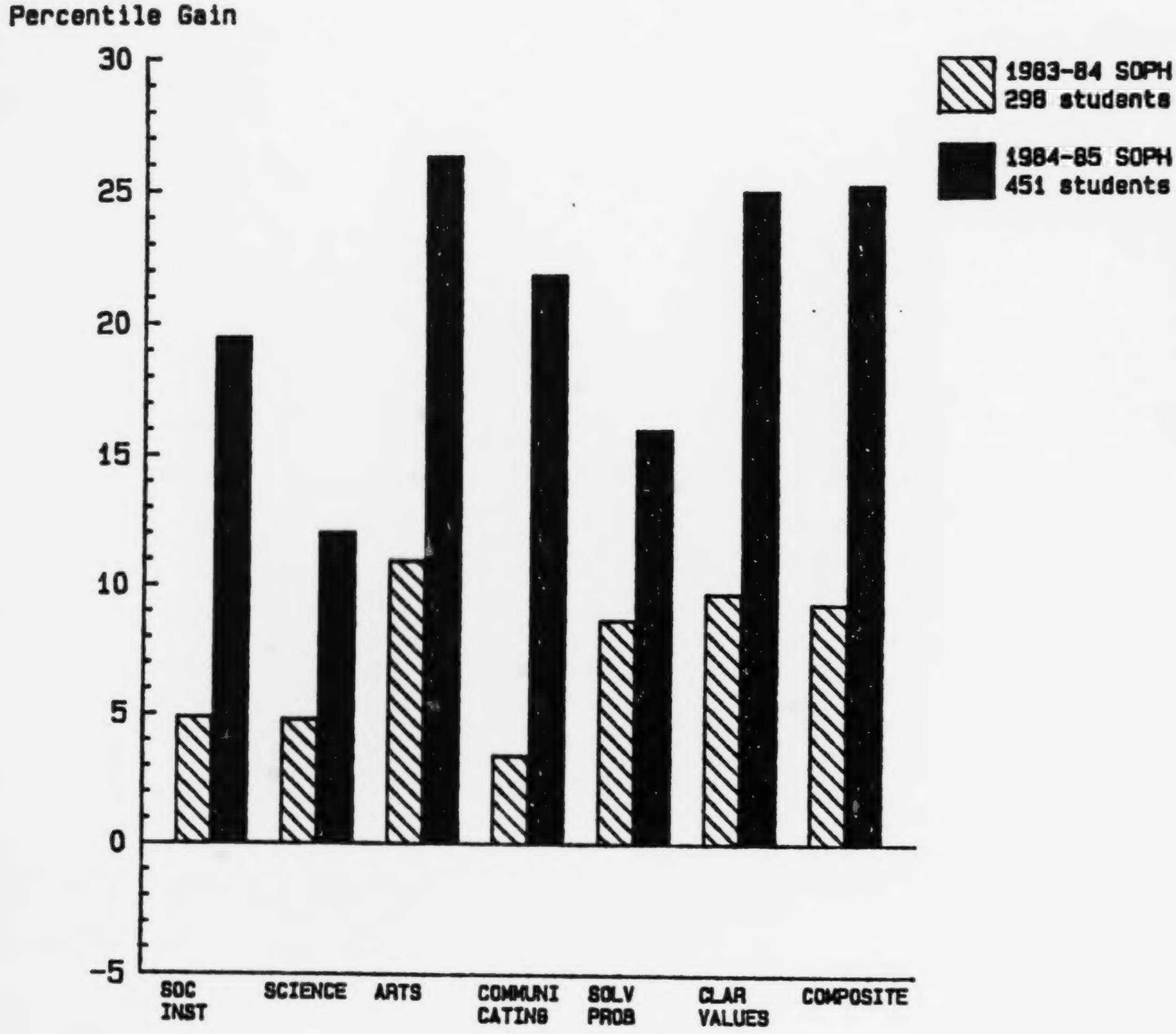


NORTHEAST MISSOURI STATE UNIVERSITY
4.0 Scale

3-5-86
---- ISS

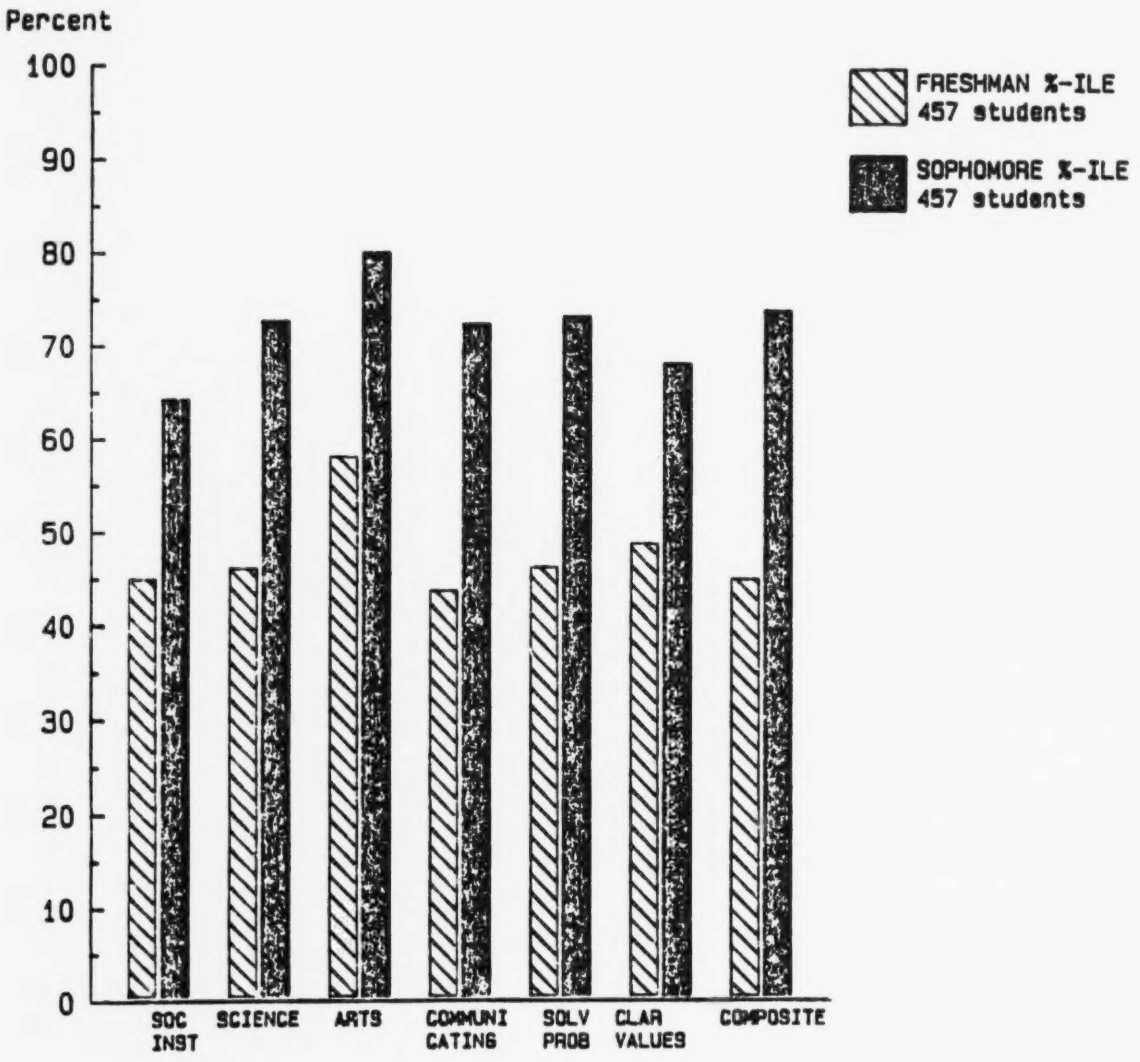
FIGURE 4.

COMP PERCENTILE CHANGE STUDY
%-ILE GAIN FRESHMAN YEAR VS. SOPHOMORE YEAR
UNIVERSITY SUMMARY



NORTHEAST MISSOURI STATE UNIVERSITY

FIGURE 5.



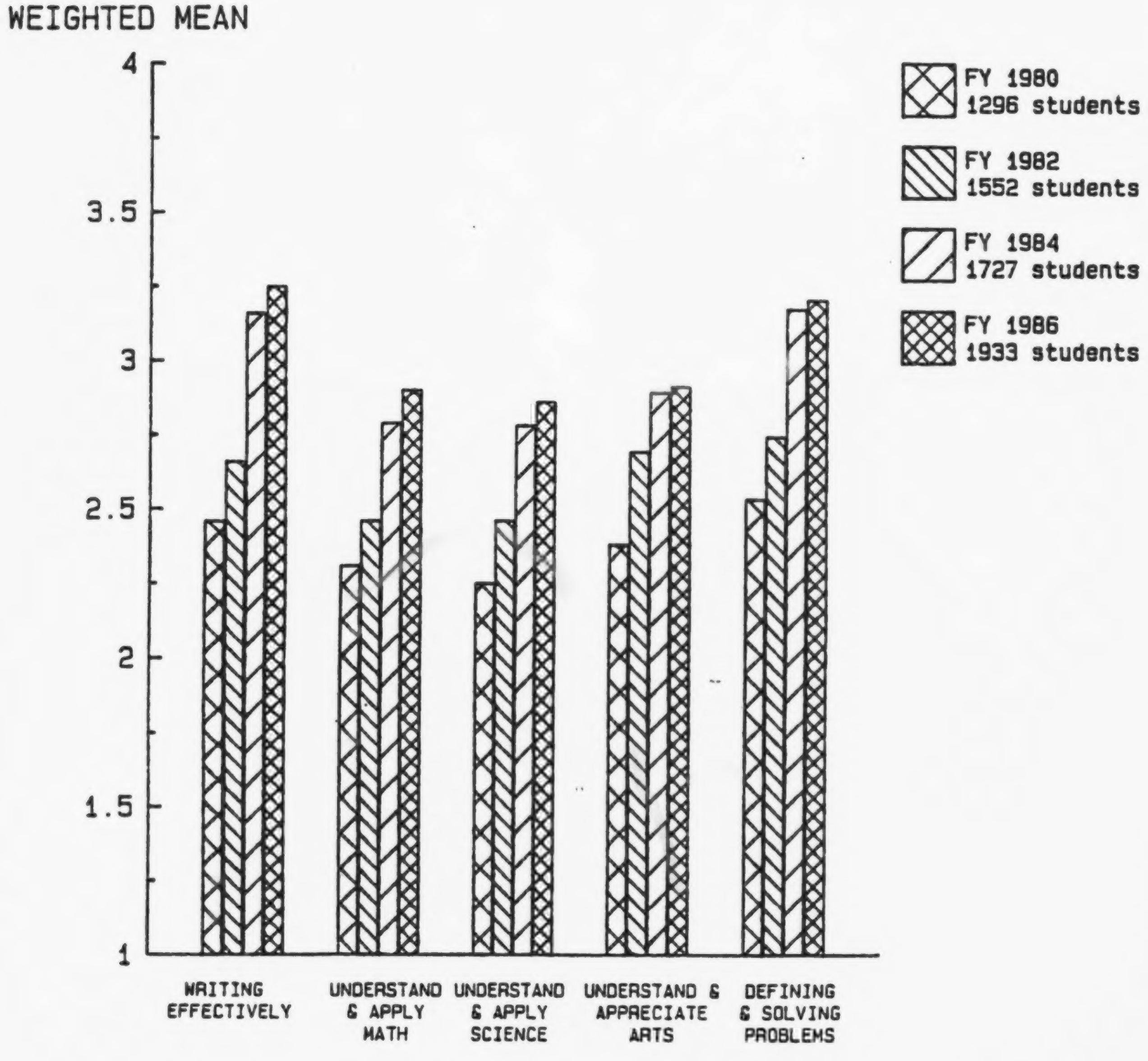
1986 COMP PERCENTILE CHANGE STUDY
University Summary
Freshman vs. Sophomore Percentile Means

NORTHEAST MISSOURI STATE UNIVERSITY

8-19-86
0007.COM

FIGURE 6.

INSTITUTIONAL STUDENT SURVEY
Weighted Means On Student Satisfaction
University Summary: Liberal Learning

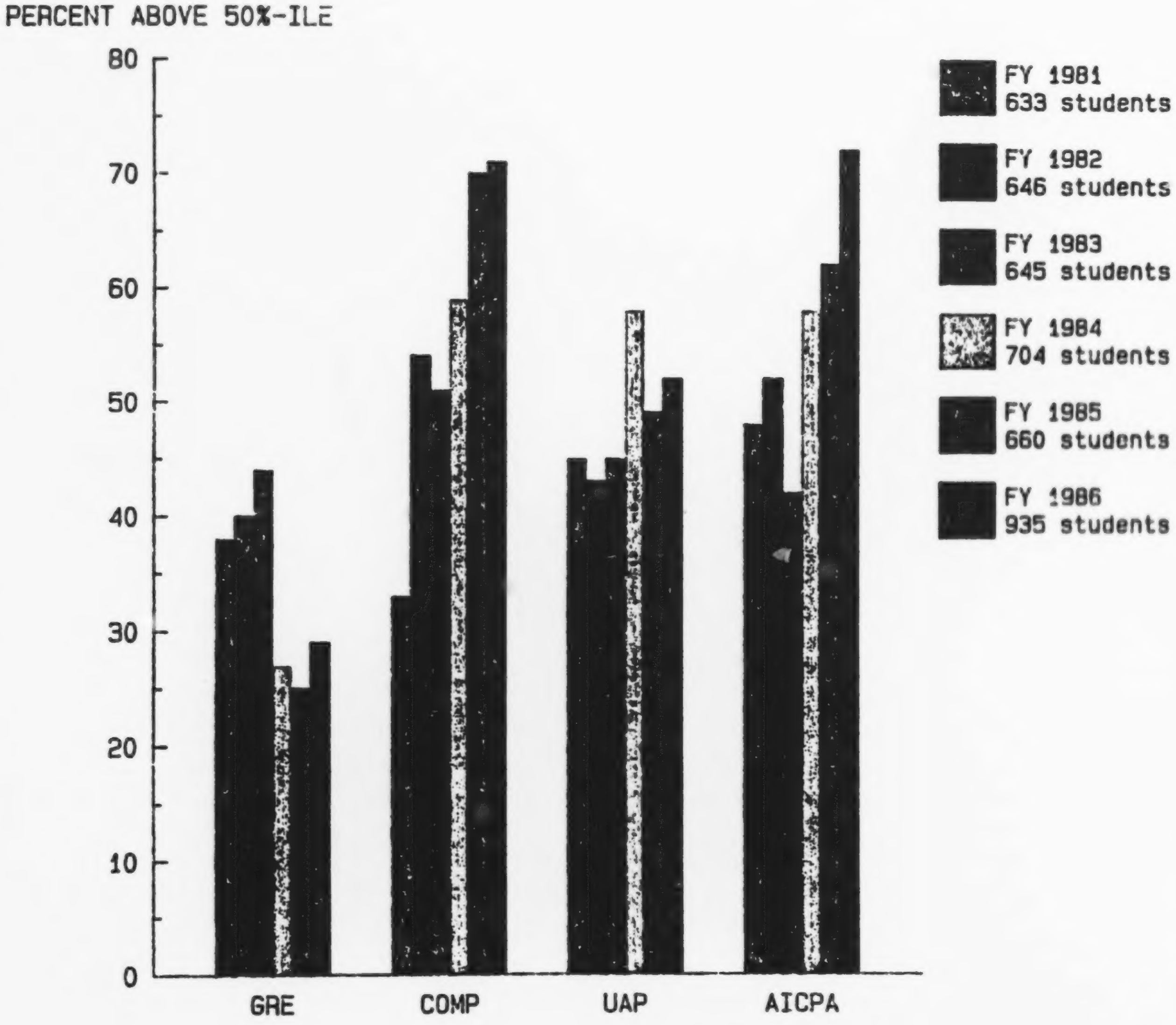


NORTHEAST MISSOURI STATE UNIVERSITY
4.0 SCALE

3-6-86
0010.ISS

FIGURE 7.

SENIOR EXAMINATION DATA
Percent Achieving 50%-ile and Above
University Summary

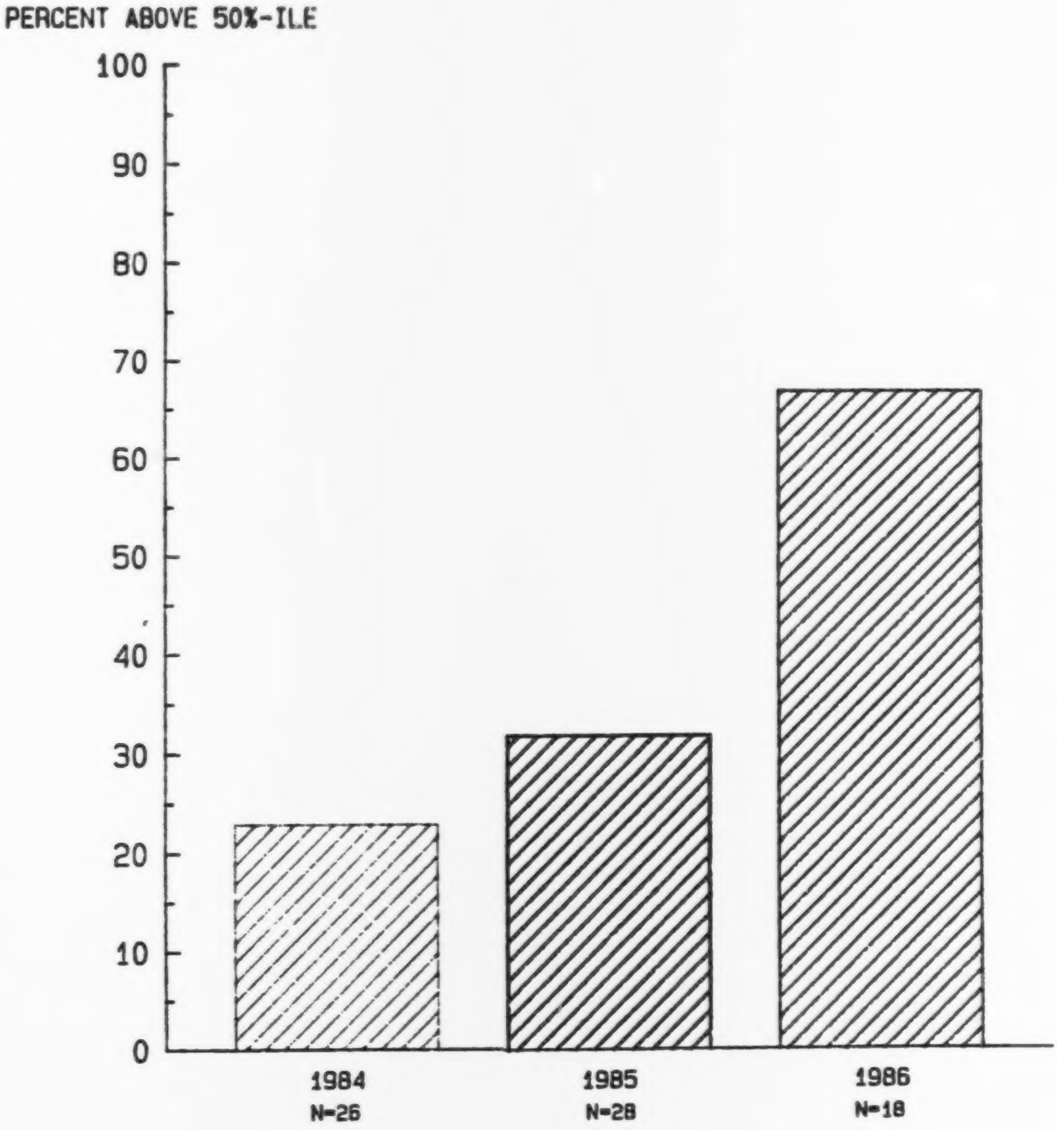


NORTHEAST MISSOURI STATE UNIVERSITY

7-14-86
0002.SEN

FIGURE 8.

GRE SENIOR EXAMINATION DATA
Percent Achieving 50%-ile and Above
Psychology Summary

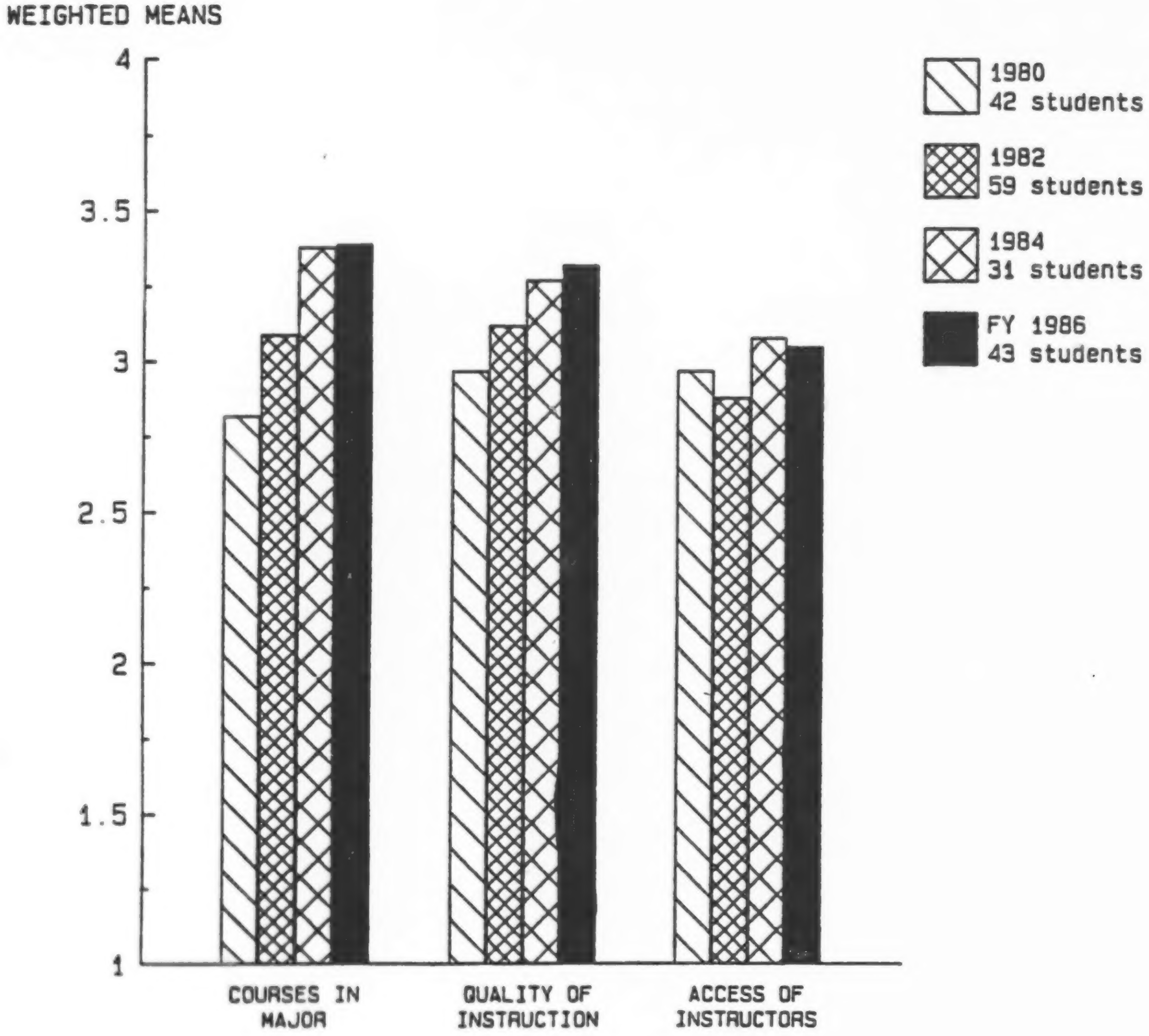


NORTHEAST MISSOURI STATE UNIVERSITY

7-14-86
0010.SEN

FIGURE 9.

INSTITUTIONAL STUDENT SURVEY
Weighted Means On Student Satisfaction
Discipline: Psychology



NORTHEAST MISSOURI STATE UNIVERSITY
4.0 Scale

3-6-86
0009.ISS

STUDENT DATA RECORD

FIGURE 10.

● THE CURRENTLY ENROLLED STUDENTS —

- Demographic Data . . . Freshman through Senior
  - — Major
  - — ACT Scores
  - — Grade-point Average
  - — Sex
  - — Age
  - — Geographic Origin
  - — Ethnicity
  - — Credit Hours
  - — Advisor
  - — Degree
  - — Senior Exam Score
  - — College Rank

- Testing Data . . . Freshman through Senior

Freshman	— ACT	(100%)
	—COMP/ACT	(50%)
Sophomore	— ACT	(50%)
	—COMP/ACT	(50%)
Junior	— no testing	
Senior	—Senior Examination in the Major	
	— GRE	
	— AICPA	
	— NTE	
	— UAP	
	— Other	

- Survey Data . . . Freshman through Senior

Entering Freshman	— Summer Orientation Student Survey
Enrolled Student	— Institutional Student Survey (distributed biennially)
Graduating Student	— Graduating Student Questionnaire
Withdrawing/ Non-returning Student	— ACT Withdrawing/Non-returning Student Survey

FIGURE 11.

INDIVIDUAL STUDENT

EXAMPLE: ACCOUNTING STUDENT '85

* **BACKGROUND**

  Rural small town environment

  Family: Business -- Butcher Shop
  -- No formal higher education

* **GOAL** uncertain: College or employment

  Parent support for college: Contingent upon serious study

  Ability: Freshman ACT Composite Score 19 (59%-ile)
  High School Rank 49/307 (84%-ile)

* **EDUCATION**

  High School: Hannibal Senior High School

  College: Enrolled at NMSU in fall 1983

  Curriculum: Liberal arts based accounting program

* **ASSESSMENT**

  FRESHMAN ACT Composite 59%-ile (Score: 19)

  SOPHOMORE ACT Composite 90%-ile (Score: 26)

  SENIOR AICPA (ACCOUNTING) 89%-ile (Score: 79)

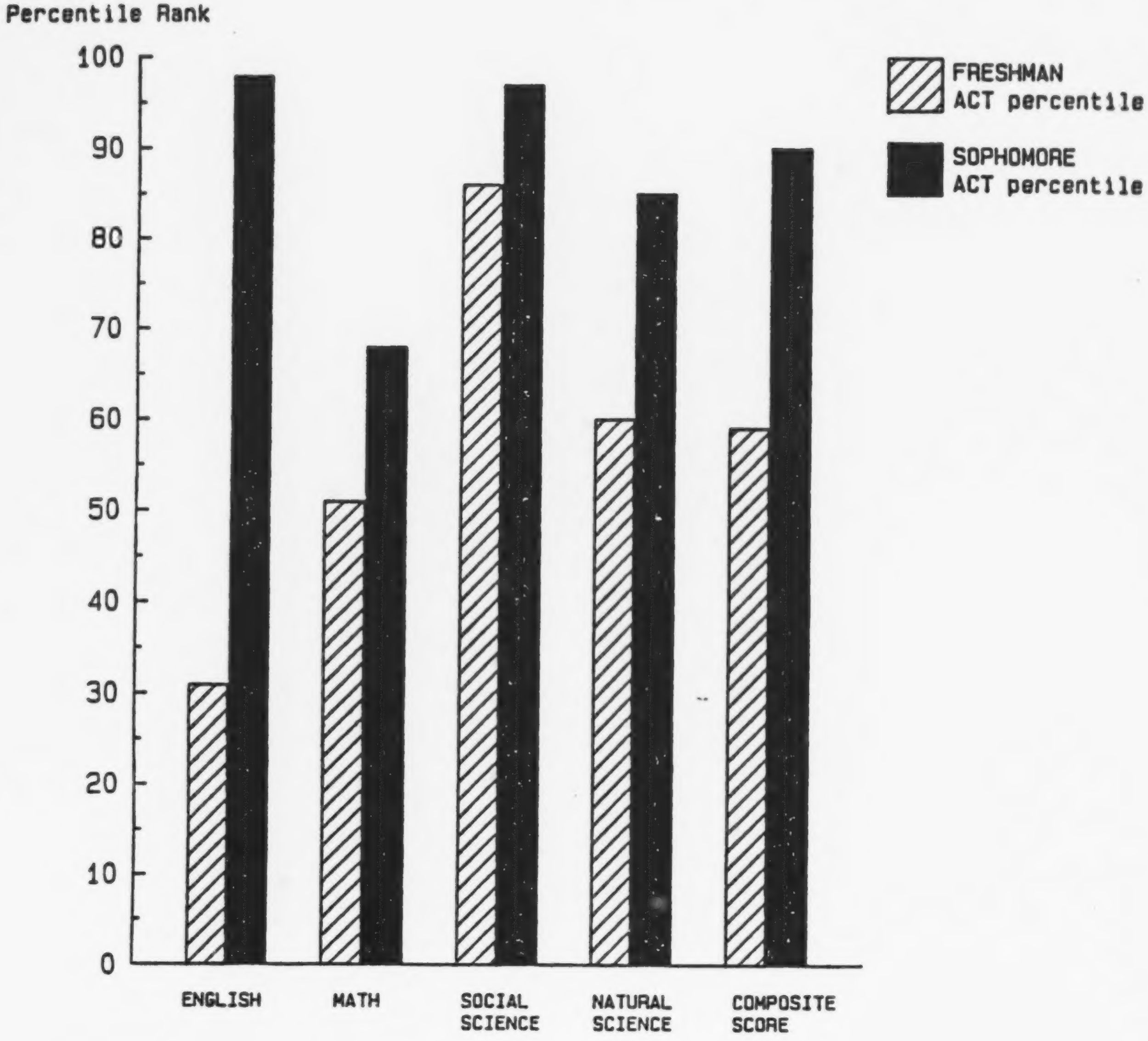
  Accum GPA - 3.72

* **GRADUATION** - B. S. in Accounting 1985

* **PLACEMENT** - Department of Revenue, Sales Tax Division
  St. Louis, Missouri

FIGURE 12.

1983 ACT PERCENTILE CHANGE STUDY
Accounting Student



NORTHEAST MISSOURI STATE UNIVERSITY

7-19-85
0013.ACT

Instructional Budget Summary

General Knowledge

INSTITUTIONAL PERFORMANCE GOAL, BUDGET FY 1984:

NMSU aims to increase the knowledge and understanding in general education as measured on the sophomore American College Testing Residual Exam (ACT) and the College Outcome Measures Project (COMP) with a four percentile-point increase in sophomore scores over freshman scores; to increase student ratings as indicated on the Graduating Student Questionnaire (GSQ) in (1) overall satisfaction toward general education by .15, and (2) cultural awareness and participation in cultural and social events by .06; to increase student ratings to 2.70 in the following as indicated on the Institutional Student Survey (ISS): (1) instruction in writing, (2) understanding and applying mathematics, (3) understanding different philosophies and cultures, (4) understanding and applying scientific principles, (5) understanding the interaction of man and his environment; and to raise satisfaction with computer services indicated on the ISS by .14.

Personal Service		1984 Program Improvements
$114,570		Five FTE Faculty
Equipment and Operations		**1984 Program Improvements**
$ 41,532	$ 3,000	Faculty Development
	9,000	Program Review
	17,500	New Personnel Costs
	9,000	Computer Software
	3,032	Films
$156,102		

Quality Performance Measurements	1982	1983	1984	w/Imprv.
SOPHOMORE TEST:				
ACT%-ile increase	1.2	2.0	2.0	4.0
COMP%ile mean increase	53%	53%	53%	57%
GSQ:				
Overall rating of General Education	2.69	2.70	2.70	2.85
Cultural awareness	2.14	2.14	2.14	2.20
ISS:				
Instruction in writing	2.66	2.66	2.66	2.70
Understanding/applying math	2.46	2.46	2.46	2.70
Understanding cultures	2.54	2.54	2.54	2.70
Understanding/applying scientific principles	2.46	2.46	2.46	2.70
Understanding interaction of man and environment	2.64	2.64	2.64	2.70
Computer services	3.01	3.01	3.01	3.15

FIGURE 13.

The addition of five FTE faculty in the area of General Education will allow the institution to place more emphasis on personalized learning as a result of reduced class size and teacher load. More individualized assistance and interest can be given to the students by the faculty. Faculty morale and preparation time will be improved. This personal concern will increase student motivation toward learning and increase satisfaction toward the general education curriculum; this in turn will affect positively knowledge gained through the comprehensive, interdisciplinary learning experiences within the general education requirement.

The monies requested for the computer-assisted instruction and other software will increase the students' awareness and knowledge of computer and other skills which are a growing requirement in good job placement for graduates.

FIGURE 14.

BUSINESS DIVISION

PERFORMANCE GOAL - BUDGET FISCAL YEAR 1985: To effect necessary gains in achievement as indicated on the Senior Exams with 56 percent scoring above the 50th percentile of the national norms; to increase the results on the Graduating Student Questionnaire (GSQ) (1) by .07 on the subject matter knowledge base, (2) by .04 on the quality of instruction, (3) by .05 on instructor accessibility, (4) by .06 on placement services, and (5) by .03 on cultural appreciation; to increase the results on the Institutional Student Survey (ISS) by .05 in the following: (1) instruction in writing, (2) understanding different philosophies and cultures, (3) academic advising, (4) value of information given by advisor, and (5) course availability.

Personal Service	
1985 Program Improvements	
$48,066	Two FTE Faculty
Equipment and Operations	
1985 Program Improvements	
$ 1,000	Faculty Enrichment/Development
1,000	Professional Services
6,000	Equipment Purchase/Replacement
7,000	New Personnel Costs
Total	
1985 Program Improvements	
$63,066	

Measurements				
Variable	1983	1984	1985	w/Imprv.
No. of Majors	1,749	1,800	1,800	
Degrees/Certificates	422	430	440	
Credit Hours	13,832	15,264	15,264	
SENIOR EXAM:				
% achieve 50%-ile	54%	54%	54%	57%
GSQ:				
Increase knowledge	2.79	2.79	2.79	2.86
Quality of instruction	3.07	3.07	3.07	3.11
Accessblty of instrctr	3.17	3.17	3.17	3.22
Placement services	3.08	3.08	3.08	3.14
Increase culture	2.08	2.08	2.08	2.11
ISS:				
Instrction in writing	2.59	2.59	2.59	2.64
Understndng cultures	2.39	2.39	2.39	2.44
Academic advising	2.88	2.88	2.88	2.93
Value of advisor info	2.83	2.83	2.83	2.88
Course availability	2.31	2.31	2.31	2.36

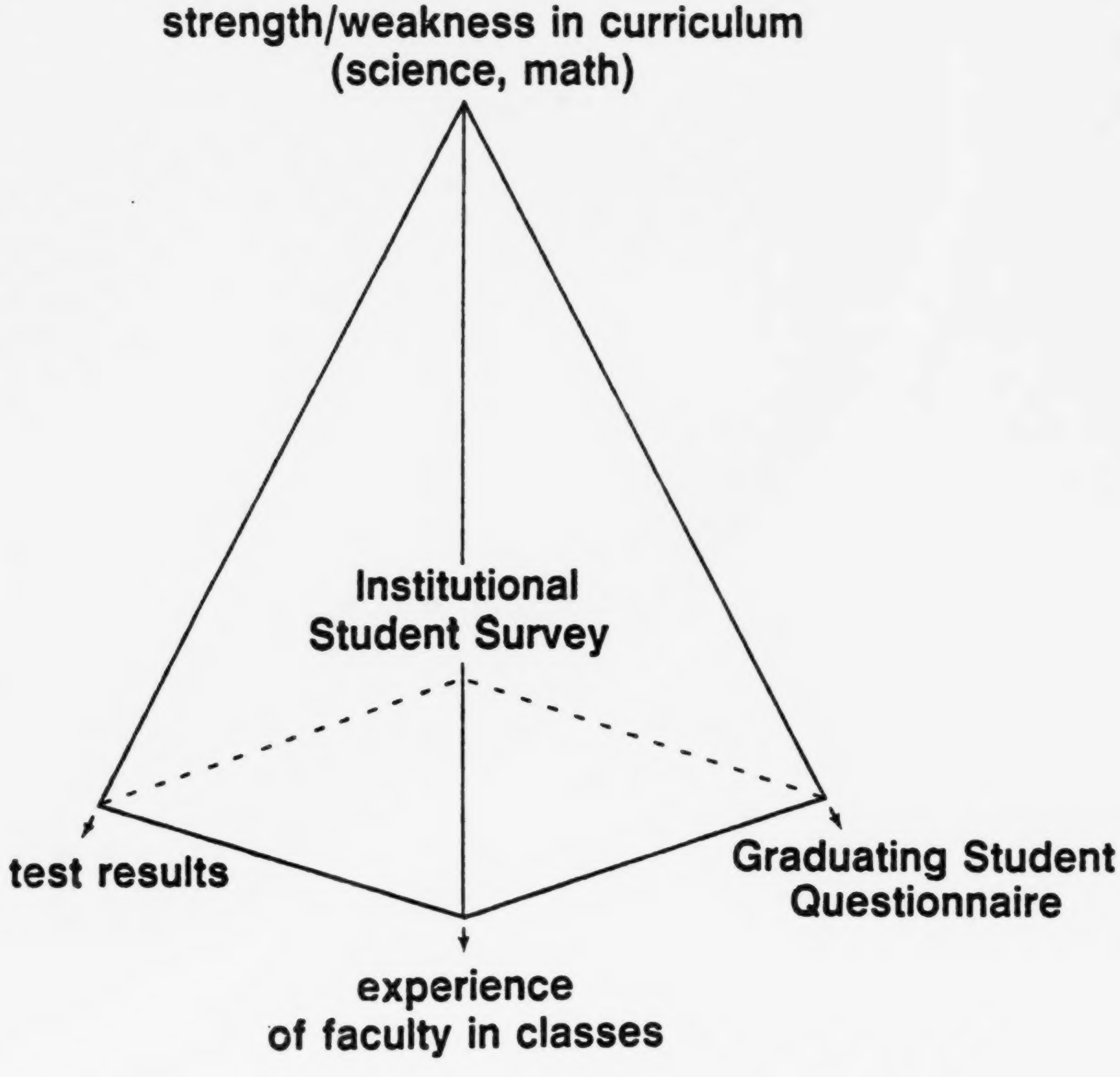
Rationale Narrative

Students enrolled in the business programs will examine business principles, theories, and policies; develop problem solving skills; and develop marketable skills to achieve success in the world of work. Faculty enrichment, professional services, and instructional supplies monies will be used to provide consultants and materials on advising, instruction, and curriculum development. Equipment replacement monies will strengthen faculty capabilities as educators as well as provide improved instructional resources to students. The employment of strong subject matter faculty and a peer counselor will help improve the knowledge base in the functional business areas. The overall effect will be greater instructor accessibility, improved quality of instruction, more involvement in placement services, increased learning, and higher achievement on the senior examinations.

FIGURE 15.

"TRIANGULATION"

(using multiple measurement techniques)



strength/weakness in curriculum
(science, math)

Institutional
Student Survey

test results

experience
of faculty in classes

Graduating Student
Questionnaire

EVALUATING ACADEMIC PROGRAMS

Hugh L. Thompson

The fundamental missions of universities have remained unchanged for more than a century. Universities provide instruction, research, and service in a wide range of areas which are constantly changing and which are delineated by custom, tradition, and regulation. Many important decisions affecting higher education are made by the Governor and the General Assembly who have the authority to enact legislation which defines the purposes of higher education and establishes certain requirements of institutional behavior to raise revenues and appropriate funds for current operations and capital improvements. Thus, public institutions must pursue missions for the greater benefit of those publics which support it. Pursuit of these missions will require responsiveness to current public needs, both state and national, as well as an intelligent anticipation of the changing needs of society.

Objectives and Assumptions

Each university has to establish institutional objectives. These objectives serve as guideposts to signal the way through circumstances which challenge the university. They cannot be considered rigid targets. Institutional objectives must assure that the university preserve and enhance its performance of these basic missions. If the campus is to meet its aspirations in these areas of primary mission, a superior faculty must be retained. The majority of funds should be allocated to those aspects of the institution which are directly involved in the provision of instruction, research, or service.

In the ensuing years, special emphasis must be placed upon faculty salaries and maintenance of key resources, such as library collections and laboratories, which are necessary for faculty productivity. The resources which provide these missions must be preserved because expenditures in these areas produce the greatest return for the publics which support the university. Such selection is the route to academic excellence.

Beginning a process of programmatic review involves agreement on certain assumptions which are likely to influence the flexibility of universities in the future. Campuses must continue to insist on high quality programs and increase their attempts to recruit outstanding students by emphasizing the quality and variety of the academic programs despite the efforts of participatory groups who may seek to protect the status quo rather than support new objectives and new organizational structures designed to improve campus efficiency and vitality. Increased emphasis on merged or system-wide activities may be desirable and institutions should target areas for excellence. Flexibility will be required to accommodate the special needs of non-traditional students and each campus must continue its strong commitment to assist students with financial need. External pressure may increase to improve efficiency by applying new technology rather than by filling vacant faculty positions. Closer ties and greater cooperation with industry should be sought by each campus.

University Groups to be Involved in Process

A university is a complex, highly decentralized organization. Faculty members, each having significant individual authority, rightly expect to be involved in major processes which will determine the future directions of an institution. Similarly, students and staff need to be informed of the processes which might lead to change, as well as the proposed changes which might emerge from the process. The institution's responsibility is to hear all voices, solicit opinion broadly, and arrive at informed, carefully-formulated policies which benefit the campus as a whole. The influence of individual committees will obviously depend not only on their breadth of perspective, but also on the persuasiveness of their evaluations and recommendations. The following list suggests the kinds of committees which are likely to be involved: Appropriate Faculty Council Committees; Graduate Council; Various Policy or Executive Committees of Schools; Council of Deans; Staff Council; *Ad hoc* internal and externa! committees; Distinguished Professors Committee; Student Advisory Committee; Board of Regents, etc.

Comparative Data

A major, but complex, step in any programmatic evaluative process is the comparison of campus top-ranked units with other top-ranked units in the same discipline or area. The purpose of the comparison is to determine in which categories top units of each campus are adequately supported or staffed and in which ones they are perceived to be deficient. The comparison should provide each campus unit with a good sense of why they are highly ranked and enable them to preserve and enhance the strengths which already exist.

Some of the information will be difficult to obtain. The comparative data will require detailed analysis and might include: number of faculty members; teaching load per FTE; average and median salary by rank; distribution of faculty effort in teaching, research, and service; age of faculty; clerical support; technical support; library allocation; equipment allocation; travel allocation; general supplies and expenses allocation; student/faculty ratio; student profiles (undergraduate and graduate); number of majors; number of credit hours; student-rated quality of teaching; student fellowship allocation; graduate and teaching assistant allocations; average and median teaching assistant salaries; placement of students; external research support per faculty memmer; faculty honors and awards (over 5-8 year period); publication record of faculty (citation index; quality of journals and presses); journals supported by unit; research centers supported by unit; size of facilities; age of facilities; uniqueness of unit; fees per student, etc.

The review process can employ both internal and external committees for each of the academic units. The quantifiable data need to be tied to qualitative judgments. The review process should be designed to determine future allocations of human, fiscal, and physical resources. In terms of service to a given geographic area of the state or to the entire state, historical and political considerations must be employed in the examination of each program.

Student Profile

In evaluating academic programs, universities need to understand the nature of their student bodies and compare their profiles with the profile of students at the universities with which they compete. As with the other comparative data, many factors need to be taken into account in assessing this information. Nevertheless, campuses should know: number of applications (freshmen, transfer, graduate); SAT/ACT/GRE distribution of applicants; high school or college rank of applicants; resident/non-resident ratio (perhaps by unit); mean GPA over time in various units; extracurricular activities of students; distribution of students by school/department/unit; housing conditions/locations/environments; distribution by sex; ethnic origins; financial aid support and prospects; attitudes of students toward the campus; recognition and awards to students; number of high school valedictorians and salutatorians enrolled; retention figures over a period of time; etc.

The list could be substantially extended. The results of the profile should help determine policies in financial aid, admissions, university relations, high school relations, career and placement services, learning support services, and academic advising. Each university should also attempt to identify the areas in which their best matriculants major to see what correlation exists between the interests of quality students and top-ranked units.

Reviews of Non-Degree Granting Units

Each campus should initiate a review of all non-degree granting units to be certain that they are operating as effectively and efficiently as possible. These reviews can be conducted in a variety of ways. However, they should be evaluated in terms of support to the essential curricular and co-curricular activities of each campus.

Establishing Priorities

After the review process has been completed, each unit should consider:

1. What existing support services or academic programs should be left unchanged?
2. What existing support services or academic programs should be modified?

3. What existing support services or academic programs should be phased out?
4. What new support services or academic programs should be offered?

Further, changes in quality or size of the unit need to be described. In regard to any changes, the priorities of the unit have to be expressed. The following information should be requested of each support service or academic program to be left unchanged, changed, or added:

1. A description of the activity.
2. How the quality of the activity is to be measured.
3. The contribution of the activity to institutional performance.
4. Enrollment projections for the activity and the basis for those projections.
5. Personnel requirements and projected expenditures.
6. Equipment needs and projected expenditures.
7. Physical facility requirements and projected expenditures.
8. Requirements for support from other operating units including library materials, support courses, admissions, counseling, other student services, and projected expenditures.
9. Internal and external constraints in the implementation or maintenance of the program.
10. Any special features of the program---its contribution to institutional and departmeatal goals, objectives, etc.

For each support service or academic program to be phased out, the following information should be requested:

1. A description of the activity.
2. How the quality of the activity is measured.
3. The contribution of the activity to institutional performance.
4. Why the activity is to be phased out.
5. The fiscal resources which will become available for other projects.

Each campus should be asked to rank their requests. The requests could be assigned to a high, medium, or low priority group. High priority requests could be implemented during the first two years of a six-year plan; medium priority items could be scheduled for implementation during the next two years. Low priority activites would not be considered until the last two years and all other requests would be scheduled for consideration in the last four years of a ten year plan.

Annual Goals and Budgeting

Each year performance goals should be established for each unit. These goals should be quantifiable, have scheduled times for completion, should be ranked in priority order, and should interface with a six and ten year plan. If a new proposl is conceived, it must coincide with the planning process and priorities which have been established. Once departmental goals are established, annual goals and priorities for the campus should be established. Again, these goals should correlate with a six and ten year plan.

Plans have to be updated with regularity, thereby enabling the campus to maintain elements of flexibility, reality, and creativity. The planning documents serve as guidelines for decision-making and should be instrumental in the preparation and construction of budgets.

Summary

Through the implementation of a comprehensive programmatic review and planning process, an institution should be able to define clearly its purposes, goals, objectives, and priorities and to outline its strategies for achieving its goals and objectives. The review and planning process should enable an institution to meld academic and financial integrity, bring into focus the resources needed to accomplish its mission programmatically, and translate to all publics, with clarity, the purpose, needs, and concerns of the campus. Foremost, a good evaluative and planning process should demonstrate to everyone that one of the best investments the state can make in maintaining and improving its vitality is through an investment in its higher educational institutions.

Hugh L. Thompson is Chancellor at the Indiana Universiiy at Kokomo, 2300 S. Washington Street, Kokomo, Indiana 46902